

QY 412 ctcttcccggggaagtactgtgacacatgtaccagctcaccagaagcccccgccac 471
Db 650 ATCTTTCCAGTGGACTCTTTCAGGAGACACATTTCCGATTACTAAACAGATGCTGG 709
QY 472 acctctgaggaagagtgtgagagcccggtgtgtctccctgggctgagttgagtc 531
Db 710 GAATTTGGAGAAAGCTTTTGGGGAAGACAAATAGTCCCTTGGAGAGACTTTGACAG 769
QY 532 ctcttgggacacctgcccacctgtgggaaccaggctgcaagccctggccttgcgcccacc 591
Db 770 GCTCTACATGAAGTGCATCCCATCAGTCTTGGGCTGGAGGCCATGGCTCTGAATCCACT 829
QY 592 attgaacctcacctgagcgaggggacgtgtctcatcttctgagtttccaggttc 651
Db 830 ATTGATCTGACCTGCAATGATTATATTTTCGGTTTGAATTTGACATCTTTACCGGACT 889
QY 652 ttccagccatggccacacctcctcaagaactggcagctcctgagcgtgcaacacccagc 711
Db 890 TITCAGCCCTGTCCTCTTTCAGGAAATGGAACAGCTTCTGTACTCTCTGCTGCTG 949
QY 712 tacatggccttctcactatgatgaggtccagagcgtctgagggcctgagggacaag 771
Db 950 TACATGGCTTTTTCAGTATGACGAAGTGAAGCTCGCTCCAGAAATTCATTCACAA 1009
QY 772 ccaggcagttacatcttcggcccaagcttactcctcctgggagcagtgggccatcggtat 831
Db 1010 CTGCGCAGTTATATCTTCGGGTGAGCTGACTCTGCTGGGTGAGTGGGCTATTGGGTAT 1069
QY 832 gtgagctcagatggcagcactctgagaccatccctgccaacaaacccctgtcccaggtg 891
Db 1070 GTTACTGTGATGGGAACATTTCTCAGACAATCCCTCACAATAAACCTCTTTCACGCA 1129
QY 892 cctctggaggagacagaagagaggtctacctctaccagatggaagaccacacaccca 951
Db 1130 CTGATGTGAGCTTCAGGGAAGGCTTCTATTGTTCTGTGATGGACGAATTCAGAACTCT 1189
QY 952 gacctgactgagctggcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1011
Db 1190 GATCTGACTGGCTTATGTGAACCACTCCCAAGACCAATCAAGTACCCAGCAACAA 1249
QY 1012 ctgagctctactgggcccactgacatctccatctgagctctgagatctgtgtgagc 1071
Db 1250 TATGAATATATATGATGAGATGGCTTCCACATTCACAACTATGTAATAATATGCTGAAAT 1309
QY 1072 aacagagtgatgaagattgagcgtgagcagcagcagcagcagcagcagcagcagcagc 1131
Db 1310 GATAAGGATGTAAGATGAGCCCTGTGGACACCTCATGTGCATCCTGTCTTACATCC 1369
QY 1132 tggcagcactcggacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1191
Db 1370 TGGCAGGAATCAGAGGTGAGGCTGCTCTTCTGCGGATGTGAATTAAGGTACTGAA 1429
QY 1192 gccgtgagtatctaccagtt 1211
Db 1430 CCATCGTGGTAGATCCGTT 1449

RESULT 2

US-08-822-586-46

; Sequence 46, Application US/08822586

; Patent No. 6015890

; GENERAL INFORMATION:

; APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSEY AND

; APPLICANT: AMALIO TELENTE

; TITLE OF INVENTION: AN EMBAC OPERON OF MYCOBACTERIA AND

; TITLE OF INVENTION: MUTANTS THEREOF

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

; STREET: 90 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,586
; FILING DATE: MARCH 20, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/437
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELETYPE: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9960
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEetical: NO
; US-08-822-586-46

Query Match 3.3%; Score 46.4; DB 5; Length 9960;

Best Local Similarity 45.2%; Pred. No. 0.047;

Matches 170; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 819 ggccatcgccgtatgtgagctcagatggcagcatcctcagaccatccctgccaacaaac 878
Db 2385 GGGCTCGGGCGCGTGCCTCGCTGGCTTCGGGCGGAACGCACTCCCTCGGATGCTC 2444
QY 879 cctgtccccgggtgctctgagggacagaagacggtcttacctctaccagatggaaa 938
Db 2445 CGCGGACCCCGCTCATGAGCAGCCGCGGACGACAACTTCGCGGACAGCATCCGGCGT 2504
QY 939 gaaccacacacagacacctgactgagctggcagcagcagcagcagcagcagcagcag 998
Db 2505 CGTCACCGGACCGAGGTGCGGACGAGGCGGACGAGCTGCGCGGCGGACGAGTCCGCGG 2564
QY 999 gtcagaggagcagctgagctctactgggcatgagctccacatttggctctgcaagt 1058
Db 2565 ATCCGCGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2624
QY 1059 ctgtctgaggaacacagagatgtaagattgagcgtgaggggacactgctctcagctg 1118
Db 2625 GCGTTCGGGACACAGCAGCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2684
QY 1119 ctgctgtgctgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagc 1178
Db 2685 CGACCGAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2744
QY 1179 caagggtcgagggcc 1194
Db 2745 CGAGGTGCGAGTGGGCC 2760

RESULT 3

US-09-320-878-23/C

; Sequence 23, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

Qy 212 ggggtcccgggggctcttggggacttttctactcatct 247
 ||| | | | | | |
Db 422 gggaagaaacatgtgtgcagcattttaacctaat 457

```

RESULT 6
PCT-US95-16542-1
/ Sequence 1, Application PC/TUS9516542
/ GENERAL INFORMATION:
/ APPLICANT: YEDA RESEARCH AND DEVELOPMENT CO. LTD.
/ APPLICANT: WEINWURZEL, Henry
/ APPLICANT: WALLACH, David
/ APPLICANT: BOLDIN, Mark
/ APPLICANT: VARFOLOMEYEV, Eugene
/ APPLICANT: METT, Igor
/ TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APOL1
/ TITLE OF INVENTION: RECEPTORS
/ NUMBER OF SEQUENCES: 2
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BROWDY AND NEIMARK
/ STREET: 419 Seventh Street N.W., Ste. 300
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: United States of America
/ ZIP: 20004
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentrin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/16542

```

	Query Match	2.8s;	Score 40;	DB 6;	Length 1701;
	Best Local Similarity	49.1s;	Pred. No. 0.88;		
	Matches 105;	Conservative 0;	Mismatches 110;	Indels 0;	Gaps 0;
Qy	32	agtggaaagagcccgccgcccctggccggggcagtcagagatgctgcagcgccctagaagac	91		
Db	242	ASGCGAAGTGGAGCGCGCTGCAGAGCGCCCTAGACCTCTTCTCATGCTGCTGGAGCAGA	301		
Qy	92	aatgcgctgaaccccggtctccgtagtcgcccttcgctgcggggacccctactgcgccgca	151		

```

Db      302   ACGACCTGGAGCCGCGGCACACCGAGACTCCTGCGCGAGCTGTCTGCCCTCCCTGCGCGGCC 361
Qy     152   cagcgacgactctcggagagtgcccatcttcggcgggcgaggcgagggccccg 211
Db      362   ACGACCTGCTTCGCGCGCGTGCAGACATCGAGAGCGGGGCGGCGCGCGCGCGCTG 421
Qy     212   ggggtcccggcggtctggggaccttttactatct 247
Db      422   GGGAGAGAGACCITGTGCGAGCAITTAACGTCATAT 457

RESULT          7
PCT-US96-10521-1
; Sequence 1, Application PC/TUS9610521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; TITLE OF INVENTION: AND OTHER PROTEINS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..768
PCT-US96-10521-1

```

[illegible]

QY 212 ggggtcccgccgctgtgggactttctactctct 247
 DB 422 GGAAGAAGACCTGTGTCAGCATTTACGTCATAT 457

RESULT 8
 US-08-916-917-1/c
 ; Sequence 1, Application US/08916917
 ; Patent No. 5856132
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len
 ; APPLICANT: Hawkins, Phillip Thomas
 ; APPLICANT: Braselmann, Sylvia
 ; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 ; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds, LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036-2811

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/916,917
 FILING DATE: 15-AUG-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/672,211
 FILING DATE: 27-JUN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Abrams, Samuel B.
 REGISTRATION NUMBER: 30,605
 REFERENCE/DOCKET NUMBER: 8549-0006-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-493-4935
 TELEFAX: 650-493-5556
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4692 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-916-917-1

Query Match 2.7%; Score 38.4; DB 3; Length 4692;
 Best Local Similarity 50.9%; Pred. No. 2.8;
 Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
 QY 4 gctctgctggtgcccctgtggggacagtggaagagcccgccctgtggcgaggca 63
 DB 1833 GCTTGGGCTGGGACAGCAGCGGAG-CCTGGGCCCGGACGGGGGAGGTGCGGGTG 1775
 QY 64 gtcaggatgctgcagcgcctagaagagcaatgcgtgcaccccccggctgtccgtgagtc 123
 DB 1774 TCCAGGGGCTGCAGAGGCTGCCGCCGACGAGGGGGCGCGCTGTCGGGCTGCC 1715
 QY 124 ccttcgtcgggagcctgtgctcccgacagcgagctgtcttcgagaggtggccattct 183
 DB 1714 TCCAGGCTGGGAGTCCCTCCGACACCATCTGGCTGGTCTTTGAAGAGTTTATAG 1655
 QY 184 cggcgccgcccggagggcccgcccgcccgcccgcccgcccgcccgcccgcccgcccg 231
 DB 1654 ATCTCTTGAACCTTCTGCCCCGGCGCGGTGGCCCCCGGCGCTCGTG 1607

RESULT 10
 US-08-916-917-1/c
 ; Sequence 1, Application US/08916917
 ; Patent No. 5856132
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len

RESULT 9
 US-08-972-631-1/c
 ; Sequence 1, Application US/08972631
 ; Patent No. 5856133
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len
 ; APPLICANT: Hawkins, Phillip T.
 ; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 ; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 2730 Sand Hill Road
 ; CITY: Menlo Park
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/972,631
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/672,211
 ; FILING DATE: 27-JUN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Halluin, Albert P.
 ; REGISTRATION NUMBER: 25,277
 ; REFERENCE/DOCKET NUMBER: 8549-0005-999
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415)854-3660
 ; TELEFAX: (415)854-3694
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4692 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: cDNA
 ; US-08-972-631-1

Query Match 2.7%; Score 38.4; DB 3; Length 4692;
 Best Local Similarity 50.9%; Pred. No. 2.8;
 Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
 QY 4 gctctgctggtgcccctgtggggacagtggaagagcccgcccgcccgcccgcccgcccg 63
 DB 1833 GCTTGGGCTGGGACAGCAGCGGAG-CCTGGGCCCGGACGGGGGAGGTGCGGGCTG 1775
 QY 64 gtcaggatgctgcagcgcctagaagagcaatgcgtgcaccccccggctgtccgtgagtc 123
 DB 1774 TCCAGGGGCTGCAGAGGCTGCCGCCGACGAGGGGGCGCGCTGTCGGGCTGCC 1715
 QY 124 ccttcgtcgggagcctgtgctcccgacagcgagctgtcttcgagaggtggccattct 183
 DB 1714 TCCAGGCTGGGAGTCCCTCCGACACCATCTGGCTGGTCTTTGAAGAGTTTATAG 1655
 QY 184 cggcgccgcccggagggcccgcccgcccgcccgcccgcccgcccgcccgcccgcccg 231
 DB 1654 ATCTCTTGAACCTTCTGCCCCGGCGCGGTGGCCCCCGGCGCTCGTG 1607

RESULT 10
 US-08-972-629-1/c
 ; Sequence 1, Application US/08972629
 ; Patent No. 5859201
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len

```

1  APPLICANT:  Hawkins, Phillip T.
2  TITLE OF INVENTION:  G-BETA-GAMMA REGULATED
3  TITLE OF INVENTION:  PHOSPHATIDYLINOSITOL-3', KINASE
4  NUMBER OF SEQUENCES:  10
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  Pennie & Edmonds
7  STREET:  2730 Sand Hill Road
8  CITY:  Menlo Park
9  STATE:  California
10 COUNTRY:  USA
11 ZIP:  94025
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  Patent In Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/972,629
20 FILING DATE:
21 CLASSIFICATION:
22
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER:  US 08/672,211
25 FILING DATE:  27-JUN-1996
26 ATTORNEY/AGENT INFORMATION:
27 NAME:  Halluin, Albert P.
28 REGISTRATION NUMBER:  25,277
29 REFERENCE/DOCKET NUMBER:  8549-0005-999
30
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE:  (415)854-3660
33 TELEFAX:  (415)854-3694
34 INFORMATION FOR SEQ ID NO:  1:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH:  4692 base pairs
37 TYPE:  nucleic acid
38 STRANDEDNESS:  single
39 TOPOLOGY:  unknown
40 MOLECULE TYPE:  CDNA
41
42 US-08-972-629-1

```

Query Match	2.7%;	Score 38.4;	DB 3;	Length 4692;
Best Local Similarity	50.9%;	Pred. No. 2.8;		
Matches 116;	Conservative 0;	Mismatches 111;	Indels 1	
QY	4	gcctcgagctgccccgctgggggagacagtgggaagagcccgccctcgggccg		
Db	1833	GCWTGGCTGGGCACGACGGGAG - CCTGGGCCGGACGGGGCAGGGTTCGG		
QY	64	gtcagatgctgcagcgctagaagagcaatcgctgaccccccgctgtccggtgag		
Db	1774	TCCAGGGGGGTGCAGAGGCTGCGGCCCCACACAGGGGCGGCCGCTCTCCGGGCT		
QY	124	ctcttcctcggggacctgctgccgcacagcgagcgtcttcagagagtgggccca		
Db	1714	TCAGGCTGCGGAGTCCCTCCGACAGCCACTCGCTGGGTGCTCTTGAAGAGTTT		
QY	184	cggcggggcggcgagagcgggcccgggggtcccgcggtctggg	231	
Db	1654	ATCCTGTGAATCTTCGCCGGCGGGGGTGGCCCCGGCGGTGTGG	1607	

RESULT 11
US-08-972-630-1/c
; Sequence 1, Application US/08972630
; Patent No. 5869271
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip T.
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3-
; NUMBER OF SEQUENCES: 10 KINASE
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds

```

, STREET: 2730 Sand Hill Road
, CITY: Menlo Park
, STATE: California
, COUNTRY: USA
, ZIP: 94025
,
, COMPUTER READABLE FORM:
,
, MEDIUM TYPE: Floppy disk
, COMPUTER: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: Patent Release #1.0, Version #1.30
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/972,630
, FILING DATE:
, CLASSIFICATION:
,
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/672,211
, FILING DATE: 27-JUN-1996
, ATTORNEY/AGENT INFORMATION:
, NAME: Halluin, Albert P.
,
, REGISTRATION NUMBER: 25,277
, REFERENCE/DOCKET NUMBER: 8549-0005-999
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (415)854-3660
, TELEFAX: (415)854-3694
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 4692 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: unknown
, MOLECULE TYPE: CDNA
,
, US-08-972-630-1

```

Query Match 2.7%; Score 38.4; DB 3; Length 4692;
Best Local Similarity 50.9%; Pred. No. 2.8;
Matches 116; Conservative 0; Mismatches 111; Indels 1

QY	4	gctctgcggctggccccctgggggagacagtgggaaggccccgcacctggggcgaggca	63
Db	1833	GCTTGGCTGGGGACAGACGGGAG-CCTGGGCCGGGACGGGCGAGGGTCGGCTG	1775
QY	64	gtcagratcttcagcgctaaagaagaatgctgacccccgcgctgcgtgagtccc	123
Db	1774	TCCAGGGGCTGCAGAGGCTCGGCCCGACGACAGGGCGGCGCGTTCGGGCTGCC	1715
QY	124	ccttcgctcgggacctgctgtcccgacacagcagctgtctcgagagggtgcccattct	183
Db	1714	TCCAGGCTCGGAGTCCCTCCGACACCATCTGGCTGGTCTCTGAAGAGTTTAG	1655
QY	184	cgggcgggccggcgaggggccccgggtccggcggtctggg	231
Db	1654	ATCCTGTTGAATCTGCGCGGGCGGGTGGGCCCGGCGCTCGTG	1607

RESULT 12
US-08-672-211-1/c
; Sequence 1, Application US/08672211
; Patent No. 5874273..

GENERAL INFORMATION:
 APPLICANT: Stephens, Len
 APPLICANT: Hawkins, Phillip T.
 TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 2730 Sand Hill Road
 CITY: Menlo Park
 STATE: California
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 18:06:44 ; Search time 152.48 Seconds
(without alignments)
3503.360 Million cell updates/sec

Title: US-09-434-708-3

Perfect score: 1422

Sequence: 1 atggctctggcgtggcccc.....ccaggaccctgcccgcc 1422

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 18783133 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36.*

1: /SID86/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID86/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID86/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID86/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID86/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID86/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID86/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID86/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID86/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID86/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID86/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID86/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID86/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID86/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID86/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID86/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID86/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID86/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID86/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID86/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID86/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	23.9	3090	20 X89079	Nucleotide sequenc
2	48	3.4	10732	20 A10594	Gene encoding a su
3	46.4	3.3	9960	19 V58939	Mycobacterium smeg
4	46	3.2	114955	20 X53491	Human adenosine Al
5	42.2	3.0	1693	21 Z56005	Contig 004 from co
6	41.6	2.9	1140	21 Z87290	S. venezuelae deso
7	41.5	2.9	3291	21 Z56004	Contig 003 from co
8	41.6	2.9	12441	21 Z87284	S. venezuelae deso
9	41.6	2.9	13613	21 Z87319	S. venezuelae deso
10	41.4	2.9	1582	20 X08910	Human FADD protein
11	41.4	2.9	1642	17 T39397	FADD (fas-associat
12	41.2	2.9	1170	20 Z39042	Human fishboy G-pr

C 13	41.2	2.9	114955	20 X53491	Human adenosine Al
C 14	40.4	2.8	880	17 T38913	B2-bradykinin rece
15	40	2.8	1701	17 T30372	MORT-1 cDNA. Homo
16	40	2.8	1701	18 T61397	MORT-1 coding sequ
17	40	2.8	1701	21 Z44745	Human FADD DNA. H
18	40	2.8	53789	19 V21187	Amcyotatopsis medi
19	39.8	2.8	627	20 V71930	MORT1 isoform MORT
C 20	39.4	2.8	1308	17 T31602	Cartilage-derived
C 21	39.2	2.8	2170	21 Z98402	Porcine betal-adre
C 22	39.2	2.8	15672	12 Q10613	Rianodin receptor
C 23	38.8	2.7	738	20 X86968	PHA acetoacetyl-Co
C 24	38.8	2.7	3632	19 V61586	Alpha-1A calcium c
25	38.8	2.7	6225	20 X52773	Human enzyme-relat
26	38.8	2.7	6225	21 A34721	Human adenosine re
C 27	38.8	2.7	6436	20 X86965	DNA containing PHA
C 28	38.8	2.7	77536	21 A14651	Nucleotide sequenc
C 29	38.4	2.7	1746	11 Q05558	Sequence encoding
30	38.4	2.7	2914	20 X57059	Human KCNO3 cDNA.
C 31	38.4	2.7	4692	19 V04633	Porcine phosphoino
C 32	38.4	2.7	4692	20 V74099	Porcine G-protein
C 33	38.4	2.7	4692	21 Z88812	Pig p101 regulator
C 34	38.2	2.7	954	19 V19862	Gallus domesticus
C 35	38.2	2.7	1347	19 V19864	Gallus domesticus
C 36	38.2	2.7	1587	19 V19863	Gallus domesticus
C 37	38.2	2.7	2233	19 V19861	Gallus domesticus
C 38	38.2	2.7	5027	19 V19860	Gallus domesticus
C 39	38	2.7	1939	11 Q04784	Sequence encoding
C 40	38	2.7	2150	20 X87820	Mouse N-acetylgluc
41	38	2.7	3318	15 Q68194	Ins5099-10 transpo
42	38	2.7	5392	15 Q64201	Sequence comprisin
C 43	38	2.7	15377	13 Q25975	MH mutant porcine
C 44	38	2.7	44377	18 T78508	Platenolide syntha
45	38	2.7	44377	18 T80414	Platenolide syntha

ALIGNMENTS

RESULT 1

X89079

ID X89079 standard; DNA; 3090 BP.

XX X89079;

XX 14-SEP-1999 (first entry)

DT Nucleotide sequence of human Cbl.

XX LAT; tyrosine kinase; linker for activation of T cell; TCR; human;
XX T-cell receptor; TCR signalling pathway; neoplasia; inflammation;
XX hypersensitivity; allergy; microbial infection; genetic disease;
XX autoimmune disease; graft rejection; modulator; Cbl; ss.

XX Homo saplens.

XX WO9932627-A2.

XX 01-JUL-1999.

XX 23-DEC-1998; 98WO-US27400.

XX 23-DEC-1997; 97US-0068690.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Samelson LE, Zhang W;

XX WPI; 1999-418925/35.

XX P-PSDB; Y27126.

XX Linker for activation of T cell protein used to, e.g. screen for
modulators of T cell signalling

PS Disclosure; Fig 12A; 125pp; English.

XX The invention relates to a protein tyrosine kinase substrate LAT (linker
CC for activation of T cells) protein. Modulation of interaction between LAT
CC and the T-cell receptor (TCR) affects the TCR signalling pathway. LAT is
CC a substrate for tyrosine kinases and becomes phosphorylated after TCR
CC engagement, resulting in recruitment of other signalling molecules. LAT
CC is used to identify and test (ant)agonists of tyrosine kinase signalling
CC pathways, i.e. modulation of interaction between tyrosine kinase
CC substrates and intracellular ligands or between these ligands and other
CC members of the pathway, including identification of downstream signalling
CC proteins, particularly in immune system cells. These modulators are
CC potentially useful as drugs and diagnostic agents, particularly for
CC diseases that involve undesirable cell proliferation, differentiation,
CC growth or T cell anergy, e.g. neoplasia, inflammation, hypersensitivity/
CC allergy, microbial infection, metabolic, genetic or autoimmune diseases,
CC graft rejection. LAT is also used to generate specific antibodies, used
CC for detection of LAT. Nucleic acid that encodes LAT, or its fragments,
CC are used to identify homologous sequences in other species; to detect the
CC LAT gene and as sources of antisense therapeutics. Modulators of LAT are
CC potentially more specific and less toxic than known immunosuppressants
CC such as cyclosporin. The present sequence represents the nucleotide
CC sequence of human Cbl.

SQ Sequence 3090 BP; 757 A; 840 C; 764 G; 729 T; 0 other;

Query Match 23.9%; Score 340; DB 20; Length 3090;
Best Local Similarity 62.2%; Pred. No. 3.3e-64;
Matches 535; Conservative 0; Mismatches 325; Indels 0; Gaps 0;

QY 352 aggcagcagctggccagctggccatctcttcagccacatgacgcagcagctggccagc 411
Db 590 agggaaacctaaaccagctgctcttcagccacatgacgcagcagctggccagc 649
QY 412 ctcttcccccgggcaagtagtactgtggagacatgtaccagctccacagggcccccgcac 471
Db 650 atcttccaaagtgcactcttcaggagacacatttcgggattactaaagcagatgctgg 709
QY 472 accttctggagggaaagtctcgagcccggtgtgtgctgcccctggctgagttgagtc 531
Db 710 gaatttggagaaagcttttgggaaagacaaatgctccctggagagcttccagag 769
QY 532 ctctgggaccctgccaccctgtggaccagctgacagccctgcccctggcgcacac 591
Db 770 gctcacatgaagtgcaccaccatcagttctgggctggagggccatggctcgaatccact 829
QY 592 attgacctaccctgaggggacgtgtccatcttcgagttcgacttctccacaggctc 651
Db 830 attgactgacctgcaatgattattctgttttgaattgacatctttaccagactc 889
QY 652 ttccagcagtgccaaacactccacagacgtgcagctcctgagcagtcacacaccagc 711
Db 890 ttccagccctggctcttctgctcaggaattggaaagccttctgtaactcactctggc 949
QY 712 tacatggctcttccactatgatgaggtccagagcgtctgcagccctgcagggagcag 771
Db 950 tacatggctttttgacgtatgacgaagtgaagctcgtccagaaattcattcacaa 1009
QY 772 ccaggcagttacatcttccggccagctgtactcgtcgtgggagtgagccatcgctat 831
Db 1010 cctggcagttatcttccggctgagctgtactcgtcgtgggctggctggtggtat 1069
QY 832 gtgagctcagatggcagctcctgcagaccatccctgcacaaacccctgctccaggtg 891
Db 1070 gtaactgtatgggaactctccagacatccctcacaaataaactcttccaaaga 1129
QY 892 ctcttgagggacagagagcgtcttaccttaccctacccagatggaaagaccacaccca 951
Db 1130 ctgatgatggcttcagggagagcgtcttcttcttcttctgagggacgaatcagatcct 1189
QY 952 gacctgactgagctcgccagggcagaaacccacagcagcgcacatcagctgtcagagagcag 1011

Db 1190 gatctgactggcttatgtgaacaaactcccaagaccatatacaaaagtacccagacaa 1249
QY 1012 ctgcagctctactggccatggcactccacatttgagctctgcagagatctgtgctgagagc 1071
Db 1250 tatgaatatactgtgagatgggtccacattccacattgtaaaatattgtgctgaaat 1309
QY 1072 aacaagagatgtaagatgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 1131
Db 1310 gataagagatgtaaaatgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 1389
QY 1132 tggcagcactggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1191
Db 1370 tggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1429
QY 1192 ggcgtgagtatctaccagtt 1211
Db 1430 cccatcgtggtagatccgtt 1449

RESULT 2
ID A10594 standard; DNA; 10732 BP.
XX A10594;
XX AC A10594;
XX DT 29-JUN-2000 (first entry)
XX DE Gene encoding a subunit of cellulose synthase.
XX KW Cellulose synthase; cellulose production; increase yield; ds.
XX OS Vigna angularis.
XX PN JP2000060568-A.
XX PD 29-FEB-2000.
XX PF 26-AUG-1998; 98JP-0239998.
XX PR 26-AUG-1998; 98JP-0239998.
XX PA (MIZUO) MIZUNO K.
XX PA (OJIP) OJI PAPER CO.
XX DR WPI; 2000-342372/30.
XX DR P-PSDB; Y85179.
XX A gene encoding a cellulose synthetic equipment - for the improvement
in the amount of cellulose synthesised in a plant body
Claim 2; Page 14-21; 32pp; Japanese.
XX This sequence represents a gene encoding a subunit of the cellulose
synthase complex of Vigna angularis. The invention relates to subunits of
cellulose synthetic equipment, that can be used to increase the amount of
cellulose synthesised by a plant. The proteins and genes encoding them
can also be used to improve the properties of the cellulose being
produced by a plant.

SQ Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match 3.4%; Score 48; DB 21; Length 10732;
Best Local Similarity 11.6%; Pred. No. 0.13;
Matches 119; Conservative 393; Mismatches 517; Indels 0; Gaps 0;
QY 257 atctggagggccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 316
Db 9654 rthrgggysgrvashashrgnthrthrgvvaasmtrmtszashsra 9753
QY 317 ccaacgacagctcttcgggggctccagactccagcagcagcagcagcagcagcagcagcagc 376
Db 9754 ssrlyssrasnthrcysshysgthrvgtytgnasvavaysthrgrgnashsua 9813

QY	377	tcattctcagccacatgcacgcagagctgcagcagcactctcccccggggcgaagtagctgtg	436
DE		Mycobacterium smegmatis embCAB operon.	
XX		Drug resistance; antibiotic resistance; antimycobacterial;	
XX		ethambutol; embCAB operon; infection; vaccine; therapy; ds.	
KW	9814	thrhsygyrasasheasyvaaraagrhvaasnashashasrsrargyrvahs	9873
XX		Mycobacterium smegmatis.	
OS	437	gacacatgacagctcaccacagcccccgcacacactctcggagggaagtgcggag	496
PN		WO9841533-A1.	
XX		24-SEP-1998.	
PD		16-MAR-1998; 98WO-US05128.	
XX		20-MAR-1997; 97US-0822586.	
PF		(YESH) UNIV YESHIVA EINSTEIN COLLEGE.	
XX		Jacobs WR, Musser J, Telenti A;	
PI		WPI; 1998-521160/44.	
XX		P-PSDB; W73055-57.	
DR		Wild type and mutated sequences of Mycobacterium embCAB operon -	
XX		useful to e.g. identify ethambutol-resistant mycobacterial strains	
PT		and produce antisense sequences to treat mycobacterial infections	
XX		Dislosure; Fig 6A-I; 62pp; English.	
PS		This is the DNA sequence of the Mycobacterium smegmatis embCAB	
XX		operon, which determines resistance to the antimycobacterial drug	
CC		ethambutol (EMB). It includes the embA, embC and embD genes that	
CC		encode proteins (see W73052-54) which are the target of action of	
CC		M. smegmatis for EMB. To identify genes conferring EMB resistance,	
CC		a genomic library from a high level EMB-resistant mutant of M.	
CC		smegmatis was introduced into wild-type M. smegmatis mc2155. Four	
CC		overlapping cosnids were identified which conferred a resistant	
CC		phenotype. The minimum size fragment capable of conferring EMB	
CC		resistance was 9 kb (pIMW99). pIMW99 plus 7 kb upstream M.	
CC		smegmatis sequence was sequenced revealing 3 homologous open	
CC		reading frames (embC, embB, embA) and 4 additional potential coding	
CC		regions. Wild-type and mutated embCAB nucleic acid sequences are	
CC		useful as probes used in the diagnosis of drug-resistant	
CC		mycobacteria or to determine the susceptibility of mycobacteria to	
CC		EMB. The nucleic acids are also useful in the treatment of	
CC		mycobacterial infections; anti-DNA or anti-RNA sequences can be	
CC		administered to inhibit embCAB operon mRNA activity (claimed). The	
CC		invention additionally provides for the use of embCAB operon	
CC		nucleic acid sequences as vaccines, or to improve existing	
CC		vaccines.	
XX		Sequence 9960 BP; 1389 A; 3592 C; 3360 G; 1618 T; 1 other;	
SQ			
Query Match 3.3%; Score 46.4; DS 19; Length 9960;			
Best Local Similarity 45.2%; Pred. No. 0.28;			
Matches 170; Conservative 0; Mismatches 206; Indels 0; Gaps 0;			
QY	819	ggccatggcgtatgtgagctcagatggcagcactctcagaccatccctgccaaacacc	878
DB	2385	ggcgtcgtgcgcgtgactcgtgggttcgggcccgaacgcatccctcgtggtctc	2444
QY	879	cctgtccaggtgctcctcggaggacagaagcagcgtctaccctaccagatggaaa	938
DB	2445	cgcggaccctcgtcagtgaggcagccggcagcagcagcagcagcagcagcagc	2504
QY	939	gaccacacacccagcactgactgagctggccagcagcagcagcagcagcagcagc	998
DB	2505	cgtcacccggcagcaggtcggcagcagcagcagcagcagcagcagcagcagcagc	2564
QY	999	gtcagaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	1058
DB	2565	atcccgcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	2624

RESULT 3
V58939
ID V58939 standard; DNA; 9960 BP.
XX
AC V58939;
XX
DT 15-FEB-1999 (first entry)


```

KW picromycin; antibiotic production; narbomycin; ds.
XX Streptomyces venezuelae.
XX OS
XX Location/Qualifiers
XX FH complement (694..1692)
XX CDS
XX /*tag= a
XX /*product= SAM_synthase
XX FT /*note= "Partial S-adenosylmethionine synthase"
XX FT complement (3..692)
XX FT /*tag= b
XX FT /*product= ORF16_product
XX FT /*note= "M. tuberculosis cbhK homologous partial protein"
XX
XX WO9961599-A2.
XX
XX 02-DEC-1999.
XX
XX 27-MAY-1999; 99WO-US11814.
XX
XX 28-MAY-1998; 98US-0087080.
XX 28-AUG-1998; 98US-0141908.
XX 22-SEP-1998; 98US-0100880.
XX 08-FEB-1999; 99US-0119139.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX PI
XX WPI; 2000-072618/06.
XX DR
XX P-PSDB; Y67216, Y67217.
XX
XX New recombinant DNA encoding a domain of narbonolide polyketide
XX synthase, for production of ketolide antibiotics -
XX
XX Disclosure; Page 38; 98pp; English.
XX

```

This is contig 004 from the recombinant cosmid pKOS023-27 DNA sequence (see Z56001) which contains a *Streptomyces venezuelae* DNA insert. The cosmid contains open reading frames which encode the various modules of the narbonolide polyketide synthase (PKS). The invention relates to the recombinant DNA containing a coding sequence for a narbonolide PKS. Polyketides are compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. Modular PKSs are responsible for the production of many antibiotics including piromycin. The narbonolide PKS consists of a loading module, six extender modules, and two thioester domains. Four proteins make up the narbonolide PKS (PICAI, PICAI1, PICAI11 and PICAIIV). PICAI includes the loading module and extender modules 1 and 2, PICAI1 includes extender modules 3 and 4, PICAI11 includes extender module 5 and PICAIIV includes extender module 6 and a type II thioesterase domain. The second type II thioesterase domain is found on the PCB protein. The nucleotide sequences encoding all of these proteins can be isolated in recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is desosaminated in *S. venezuelae* to yield narbomycin, the desosamyl transferase enzyme is required for this conversion, and the desosamine biosynthetic genes are also found in cosmid pKOS023-27. The recombinant DNA of the invention is used to express, in transformed cells, narbonolide (or its derivatives) or other ketolides (particularly hybrids), which may then be converted (e.g. by other enzymes recombinantly expressed in the same hosts) to polyketide antibiotics or their intermediates. The antibiotics are useful in human or veterinary medicine.

Sequence 1693 BP: 237 A; 566 C; 633 G; 257 T; 0 other;

	Query Match	3.08;	Score 42.2;	DB 21;	Length 1593;
	Best local Similarity	48.9%	Pred. No. 1.6;		
	Matches 113;	Conservative	0;	Mismatches 118;	Indels 0;
Qy	797	gctctactcgcctggggcagtgggccatcggctctatggagctcagatggcagcatccctgc	856		
Db	971	gctcgcagagtcctcaggttcgcgtacgcgattgcgcgaagccgcagcccgctcttcctgcgc	912		

857 agaccatcctgcacacaacccctgtcccagtgctcctcgagggaacagaagcgctt 916
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 AGACCTTCGGCACCCACAGATCGAGACCAGAGAAGATCGAAGACGCCATCGGCGAGTCT 852

Qy 917 tctacctaccacgatgaaacaccccccaacccacacccagtactgagctggcgagcgag 976
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 851 TCGACCTCGCGCCGCGCGATCATCCGCGACCTCGACCTGCCTGCTCCTCGCGATCTACTCCC 792

Qy 977 aaoccagagcgcatccaogttcagagggagcagctgcagctctactggg 1027
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 791 AGACCGCGCGCTACGGCCCACTTCGGCGCGAGCTGCGGGACTTCACTGTGG 741

RESULT 6
287290/C
ID 287290 standard; DNA; 1140 BP.
XX
XX 287290;
XX
XX DT 05-JUN-2000 (first entry)
XX
S. venezuelae desosamine biosynthetic gene desv, SEQ ID NO:15.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymers; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolemia; crop protection agent; ds.
XX
XX Streptomyces venezuelae ATCC15439.
XX
XX Key Location/Qualifiers
FH FT CDS 1..1140
FT FT /*tag= a
FT FT /product= "Desv"
PN WC2000000620-A2.
XX
XX PD 06-JAN-2000.
XX XX 25-JUN-1999; 99WO-US14398.
XX XX 26-JUN-1998; 98US-0105537.
XX PA (MINU) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
PI WPI; 2000-160679/14.
XX DR P-PSDB; Y77185.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin -
XX
XX Claim 3; Page 362; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomycetes antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to

CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. Sequences 287286-287294 represent desosamine
 CC biosynthetic genes from Streptomyces venezuelae ATCC 15439, which encode
 CC proteins Y77181-Y77189.

XX Sequence 1140 BP; 163 A; 472 C; 373 G; 132 T; 0 other;

Query Match 2.9%; Score 41.6; DB 21; Length 1140;
 Best Local Similarity 46.3%; Pred. No. 2.1;
 Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 73 ctacagcgctagaagcaatcgctgaccccccgcgtgctccgtgagtcctcccttcgctg 132
 DB 1125 CGCTCGGCCCATTCGGCAGCGGTCGATCACCCGACGGCTCGGGCGGCTCCAGGTG 1066
 QY 133 cgggacctgctgcccgcacagcgacgtcttcgagaggtggcccatcttcgggggag 192
 DB 1065 CGGGCGGATCGGCAGGCTGAGGACCTGCGCGGAGACTCTCGGCCCGCGGAGCGGCC 1006
 QY 193 gcggcgagagcgcccccgggggtcccgcggtctggtgggaattttactatctacctg 252
 DB 1005 TTCGGCGGTCGCTCCCGCGGTAGCGCGGCGAGAGGTGACGGGTAGTGGCT 946
 QY 253 gccaattcgagggccaagacgagcgaggtggcgcgctgctccctcccgggcggaag 312
 DB 945 GAGGGTGTGATCCCGGGGCTCGAGGTGGTGGCACCTGTGCGCGGCTCGGTGCG 886
 QY 313 agtgcacaagacagatcttctcggggcggtgtccagactcagcgacagctggccaa 368
 DB 885 CACGGTGAAGAGGTGCACAGCGGGTCGGTGTGCGGCGCGGTACCGGAGGCGCA 830

RESULT 7

256004/c
 ID 256004 standard; DNA: 3291 BP.

XX AC 256004;

XX DT 23-MAR-2000 (first entry)

DE Contig 003 from cosmid pKOS023-27 from Streptomyces venezuelae.

XX Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 003;
 KW ketolide; dNDP glucose synthase; dehydratase; picromycin; narbomycin;
 KW antibiotic production; ds.

XX Streptomyces venezuelae.

XX Key Location/Qualifiers
 FT CDS 104..982
 FT /*tag= a
 FT /product= dNDP_glucose_synthase
 FT /note= "glucose-1-phosphate thymidyl transferase"
 FT CDS 1114..2127
 FT /*tag= b
 FT /product= dNDP_glucose4_6_dehydratase
 FT 2124..3263
 FT /*tag= c
 FT /product= PICCI
 FT /transl_except= (Pos:2874..2876, aa:Xaa)
 FT /note= "Xaa = Unknown"

XX WO961599-A2.

XX 02-DEC-1999.

XX 27-MAY-1999; 99WO-US11814.
 XX 28-MAY-1998; 98US-0087080.
 PR 28-AUG-1998; 98US-0141908.
 PR 22-SEP-1998; 98US-0100880.
 PR 08-FEB-1999; 99US-0119139.
 XX (KOSA-) KOSAN BIOSCIENCES INC.
 XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
 XX WPI; 2000-072618/06.
 DR P-PSDB; Y67206, Y67214, Y67215.
 XX New recombinant DNA encoding a domain of narbonolide polyketide
 PT synthase, for production of ketolide antibiotics -
 XX Disclosure; Page 37-38; 98pp; English.
 XX This is contig 003 from the recombinant cosmid pKOS023-27 DNA sequence
 CC (see 256001) which contains a Streptomyces venezuelae DNA insert. The
 CC cosmid contains open reading frames which encode the various modules of
 CC the narbonolide polyketide synthase (PKS). The invention relates to
 CC recombinant DNA containing a coding sequence for a narbonolide PKS.
 CC Polyketides are compounds synthesised from 2-carbon units through a
 CC series of condensations and subsequent modifications. Modular PKSs are
 CC responsible for the production of many antibiotics including picromycin.
 CC The narbonolide PKS consists of a loading module, six extender modules,
 CC and two thioester domains. Four proteins make up the narbonolide PKS
 CC (PICAI, PICAILI, PICAIII and PICAIIV). PICAI includes the loading module
 CC and extender modules 1 and 2, PICAILI includes extender modules 3 and 4,
 CC PICAIII includes extender module 5 and PICAIIV includes extender module 6
 CC and a type II thioesterase domain. The second type II thioesterase domain
 CC is found on the PICB protein. The nucleotide sequences encoding all of
 CC these proteins can be isolated in recombinant form from the recombinant
 CC cosmid pKOS023-27. Narbonolide is desosaminylated in S. venezuelae to
 CC yield narbomycin, the desosaminyl transferase enzyme is required for this
 CC conversion, and the desosamine biosynthetic genes are also found in
 CC cosmid pKOS023-27. The recombinant DNA of the invention is used to
 CC express, in transformed cells, narbonolide (or its derivatives) or other
 CC ketolides (particularly hybrids), which may then be converted (e.g. by
 CC other enzymes recombinantly expressed in the same hosts) to polyketide
 CC antibiotics or their intermediates. The antibiotics are useful in human
 CC or veterinary medicine.
 XX SQ Sequence 3291 BP; 521 A; 1289 C; 1044 G; 436 T; 1 other;

Query Match 2.9%; Score 41.6; DB 21; Length 3291;
 Best Local Similarity 46.3%; Pred. No. 2.5;
 Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 73 ctgcagcgctagaagcaatcgctgaccccccgcgtgctccgtgagtcctcccttcgctg 132
 DB 3247 CGCTCGGCCCATTCGGCAGCGGTCGATCACCCGACGGCTCGGGCGGCTCCAGGTG 3188
 QY 133 cgggaacctgctgcccgcacagcgagctgttcgagaggtggcccatcttcggggcg 192
 DB 3187 CGGGCCGATCGGCAGGCTGAGGACCTGCGCGGCGAGCTCTCGGCCCGGGAGCGAGCC 3128
 QY 193 gcggcgagagcgcccccgggggtcccgcggtctcgggaattttactatctacctg 252
 DB 3127 TTCGGCGGTCGCTCCCGCGGTAGCGCGGCGAGAGGTGACCGGGTAGTGGCT 3068
 QY 253 gccaattcgagggccaagacgagcgaggtggcgcgctgctccctcccgggcggaag 312
 DB 3067 GAGGGTGTGATCCCGGGGCTCGAGGTGGTGGCACCTGTGCGCGGCTCGGTGCG 3008
 QY 313 agtgcacaagacagatcttccggcggtgtccagactcagcgacagctggccaa 368
 DB 3007 CACGGTGAAGAGGTGCACAGCGGGTCGGTGTGCGGCGCGGTACCGGAGGCGCA 2952

FT		/*tag= i	
FT		/product= "PiKB gene cluster protein #9 (Y77210)"	
FT	CDS	complement (13706..15043)	
FT		/*tag= j	
FT		/product= "PiKB gene cluster protein #10 (Y77211)"	
FT	CDS	15404..15574	
FT		/*tag= k	
FT		/product= "PiKB gene cluster protein #11 (Y77212)"	
XX			
XX	WO200000620-A2.		
XX			
XX	06-JAN-2000.		
XX			
XX	25-JUN-1999;	99WO-US14398.	
XX			
XX	26-JUN-1998;	98US-O105537.	
XX		(MINU) UNIV MINNESOTA.	
PA	Sherman DH, Liu H, Xue Y, Zhao L;		
PI	WPI: 2000-160679/14.		
DR	P-FSDB; Y77204, Y77205, Y77206, Y77207, Y77207, Y77208, Y77209,		
DR	Y77210, Y77211, Y77212, Y80998, Y80999.		
XX			
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.		
PT	synthesis of methymycin and pikromycin -		
XX	Disclosure: Figure 32; 438pp; English.		
XX			
CC	The invention relates to an isolated and purified nucleic acid segment		
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its		
CC	biologically active variant, where the nucleic acid sequence is not		
CC	derived from the eryC gene cluster of Saccharopolyspora erythraea or		
CC	Streptomyces antibioticus. The invention also relates to a macrolide		
CC	biosynthetic gene cluster, or fragments thereof. The macrolide		
CC	biosynthetic gene cluster encodes proteins which synthesise methymycin,		
CC	pikromycin, neomethymycin, narbomycin or a combination of these		
CC	compounds. Recombinant or augmented cells comprising the desosamine		
CC	and/or macrolide biosynthetic gene clusters are useful for the production		
CC	of biologically active macrolides. The macrolide biosynthetic proteins		
CC	are useful for synthesis of methymycin, pikromycin, neomethymycin and		
CC	narbomycin. The alternative termination of polyketide synthesis may be		
CC	useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)		
CC	monomers. The compounds produced by the recombinant host cells are useful		
CC	as biopolymers, e.g., in packaging or biomedical applications, to		
CC	engineer PHA monomer syntheses or to prepare biologically active agents,		
CC	such as chemotherapeutics, immunosuppressants, agents to treat asthma,		
CC	chronic obstructive pulmonary disease as well as other diseases involving		
CC	respiratory inflammation, cholesterol-lowering agents or macrolide-based		
CC	antibiotics which are active against a variety of organisms, e.g.,		
CC	bacteria, including multi-drug resistant pneumococci and other		
CC	respiratory pathogens, as well as viral parasitic pathogens, or as crop		
CC	protection agents (e.g., fungicides or insecticides) via expression of		
CC	polyketides in plants. The present sequence represents the desosamine		
CC	biosynthetic gene cluster (pix) from Streptomyces venezuelae ATCC		
CC	15439, as given in figure 32.		
XX			
SQ	Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other:		
	Query Match	2.9%; Score 41.6; DB 21; Length 13613;	
	Best Local Similarity	46.3%; Freq. No. 3.2;	
	Matches 137; Conservative	0; Mismatches 159; Indels 0; Gaps	
QY	73 ctgcgagcgtctagaagacgaatgcctcaccccggtgtccgtgagtcccccttcgctg 132		
Dd			
	6845 cgcctcgcccatcgcgaggcggtgatcaccgcgagcctgcggcgctccagggtg 6904		
QY	133 cgggacctgctcccgcacagcgagctgtcttcgagagtgggccattctcggcgggcg 192		
Dd			
	-6905 cgggcgcatcgcgagctgaggacctgcgcgcaagctctcggcccggggagcgacc 6964		

* CC probes that identify wild-type or mutant coding sequences. They are
CC also useful for expressing FIP mRNA, proteins or fusion proteins,
CC and in the generation of FIP antisense oligonucleotides and
CC ribozymes. They are also useful in expression constructs and in
CC gene delivery vehicles (optionally in combination with a condensing
CC agent) that deliver FIP mRNA or oligonucleotides, FIP proteins
CC (including variants), FIP-specific ribozymes or single-chain
CC antibodies into eukaryotic cells. This is the human FADD protein
CC gene. Human FIP protein binds to amino acids 1-110 of the human
CC FADD protein given in W96134.
XX
SQ Sequence 1582 BP; 344 A; 433 C; 483 G; 322 T; 0 other;

Query Match 2.9%; Score 41.4; DB 20; Length 1582;
Best Local Similarity 48.9%; Pred.No. 2.4; Indels 0; Gaps 0;
Matches 111; Conservative 0; Mismatches 116;
QY 21 gtggggcgacagtgggaagagccgcgccttggccggcagtcagatgctgcagcg 80
Db 156 ggcggtgggcaagcgaagctgagcgctgcagagcgcttagacctttctcatgct 215
QY 81 cctagaagacatgctgcagcccgctgctgcagtcagtccttcctgcgggaacct 140
Db 216 gctggagcagaagacctgagcccgccacaccagctcctgcgcgagctgctgcctc 275
QY 141 gctgcccgcacagcagcagctgcttcagagagtggtgcccattctcggcgccgcgg 200
Db 276 cctgcggccacagacctgctgcggcgctgcagacctcgagggcgccggcgccgg 335
QY 201 aggcggcccggggtcccgcgctctctgggaactttctactatct 247
Db 336 ggcgcgcctgtgggaagaaagacctgtgtgcagcatttaacgtcatat 382

RESULT 11
T39397
ID T39397 standard; cDNA; 1642 BP.
AC T39397;
DT 22-FEB-1997 (first entry)
DE FADD (Fas-associating protein with novel death domain) gene.
KW Human; FADD; Fas-associating protein with novel death domain;
KW apoptosis; Fas receptor; death domain; gene therapy; antibody;
KW immunoassay; drug screening; diagnostic; AIDS; antiinflammatory;
KW antitumour; cerebroprotective; neuroprotective; ss.
OS Homo sapiens.

XX Key Location/Qualifiers
FH 5'UTR 1..129
FT /tag= a
FT misc_feature 4..6
FT /tag= b
FT /note= "In-frame stop codon"
FT CDS 130..756
FT /tag= c
FT /product= FADD protein
FT misc_feature 133..501
FT /tag= d
FT /note= "Encodes N-terminal half, inducing apoptosis
FT but not binding Fas receptor"
FT misc_feature 198
FT /tag= e
FT /note= "Clone 15 start point"
FT misc_feature 249
FT /tag= f
FT /note= "Clone 8 start point"
FT misc_feature 460..860
FT /tag= g

FT 3'UTR /note= "Region encoding death domain"
FT 757..1642
FT /tag= h
FT polyA_signal 1636..1641
FT /tag= i
XX W09631603-A2.
XX 10-OCT-1996.
XX 28-FEB-1996; 96WO-US02857.
XX 18-MAY-1995; 95US-0443982.
XX 03-APR-1995; 95US-0416379.
XX (UNMI) UNIV MICHIGAN.
XX Dixit VM, O'Rourke K;
XX WP1; 1996-465026/46.
XX P-PSDB; W03653.

FADD protein that binds to cytoplasmic region of Fas receptor - for
identifying inhibitors of Fas-associated apoptosis useful for
treating e.g. AIDS, leukaemia, stroke, etc
Example 1; Fig 2A-B; 96pp; English.

The sequence encodes FADD (Fas-associating protein with novel death
domain), which binds the cytoplasmic region of a Fas receptor, and
modulates apoptosis induced by activation of the receptor by ligand
binding. The cDNA has been isolated using a yeast two-hybrid system
which screens for proteins interacting with the Fas cytoplasmic
domain. A GAL4 DNA-binding domain has been fused to the human Fas
antigen cytoplasmic tail to form a bait plasmid, which is used with
a prey plasmid, containing a human B-lymphocyte cDNA library fused to
the GAL4-activation domain, to co-transform yeast cells.
Overlapping clones 8 and 15 have been isolated, and the full-length
cDNA has been isolated from a human umbilical vein endothelial cell
(HUVEC) library using clone 15 as a probe. The cDNA has an
in-frame stop codon 130 bp upstream of the initiator Met. The
encoded protein contains a death domain, with interactions with the
death domain of Fas. The DNA may be used in gene therapy, and the
protein or a corresponding antibody may be used to screen for
agents modulating FADD pathway cellular functions and Fas-associated
apoptosis, for use in therapy of e.g. AIDS, inflammation, etc.
leukaemia, myocardial infarction, degenerative disease, etc.
SQ Sequence 1642 BP; 354 A; 448 C; 508 G; 332 T; 0 other;

Query Match 2.9%; Score 41.4; DB 17; Length 1642;
Best Local Similarity 48.9%; Pred.No. 2.4; Indels 0; Gaps 0;
Matches 111; Conservative 0; Mismatches 116;
QY 21 gtggggcgacagtgggaagagccgcgccttggccggcagtcagatgctgcagcg 80
Db 216 ggcggtgggcaagcgaagctgagcgctgcagagcgcttagacctttctcatgct 275
QY 81 cctagaagacatgctgcagcccgctgctgcagtcagtccttcctgcgggaacct 140
Db 276 gctggagcagaagacctgagcccgccacaccagctcctgcgcgagctgctgcctc 335
QY 141 gctgcccgcacagcagcagctgcttcagagagtggtgcccattctcggcgccgcgg 200
Db 336 cctgcggccacagacctgctgcggcgctgcagacctcgagggcgccggcgccgg 395
QY 201 aggcggcccggggtcccgcgctctctgggaactttctactatct 247
Db 396 ggcgcgcctgtgggaagaaagacctgtgtgcagcatttaacgtcatat 442

RESULT 12

Z39042/c
 ID Z39042 standard; cDNA; 1170 BP.
 AC Z39042;
 XX
 DT 25-FEB-2000 (first entry)
 XX
 DE Human fishboy G-protein coupled receptor encoding cDNA.
 XX
 KW Human; fishboy; G-protein coupled receptor; therapy; diagnosis;
 KW signal transduction; HIV; infection; pain; cancer; diabetes; obesity;
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;
 KW hypotension; hypertension; urinary retention; stroke; osteoporosis;
 KW angina pectoris; myocardial infarction; migraine; ulcer; allergy;
 KW benign prostatic hypertrophy; vomiting; dyskinesia; psychotic disorder;
 KW neurological disorder; Huntington's disease; Tourette's syndrome; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1170
 FT /*tag= a
 FT /product= "Fishboy"
 FT /note= "a G-protein coupled receptor"
 XX
 XX W09942484-A1.
 XX
 XX 26-AUG-1999.
 XX
 XX 16-FEB-1999; 99WO-US03246.
 XX
 XX 20-FEB-1998; 98US-0075626.
 XX 13-AUG-1998; 98US-0133132.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 XX Elshourbagy N, Lane PA, Tsui P, Vawter L;
 XX
 XX WPI; 1999-527456/44.
 XX P-PSDB; Y57456.
 XX
 XX A new G-protein coupled receptor, Fishboy used to treat, e.g. benign
 XX prostatic hypertrophy -
 XX
 XX Claim 2; Page 34-35; 49pp; English.
 XX
 XX The present sequence encodes a human G-protein coupled receptor,
 XX designated Fishboy. Fishboy products can be used to treat subjects in
 XX need of modulated Fishboy expression or activity. Diseases or disorders
 XX to be treated include bacterial, protozoal, fungal and viral,
 XX particularly HIV-1 or HIV-2 infections, pain, cancers, diabetes, obesity,
 XX anorexia, bulimia, asthma, Parkinson's disease, acute heart failure,
 XX hypotension, hypertension, urinary retention, osteoporosis, angina
 XX pectoris, myocardial infarction, stroke, ulcers, allergies, benign
 XX prostatic hypertrophy, migraine, vomiting, psychotic and neurological
 XX disorders and dyskinesias such as Huntington's disease or Tourette's
 XX syndrome.
 XX
 XX Sequence 1170 BP; 137 A; 409 C; 415 G; 209 T; 0 other;
 SQ

Query Match 2.9%; Score 41.2; DB 20; Length 1170;
 Best Local Similarity 44.3%; Pred. No. 2.5;
 Matches 210; Conservative 0; Mismatches 263; Indels 1; Gaps 1;
 QY 8 tggcggtggccgctggggcgacagtgggaagagccgcgcctgggcccggcgagta 67
 DB 815 TGGTAGGGGGCCGAGCAAGCCGAGGACGACGATGGCGCTACACAGCGCGCCACC 756
 QY 68 ggatgctgcagcgcctagaagacaaagcgtcgaccccggtgctgctgagtcctccctt 127
 DB 755 CGCGCCCGTGGCGCGGAGCCCGACGCGGCGCCCGCGAGCGTCCAGCGTACGCTG 696

QY 128 cgctggggaactgctgcccgcacagcgagctgcttcgagagtggtggccattctc-gg 186
 DB 695 TAGCAGCCGAGCATCAGCCGAAAGACGAAAGCGGTTCAGAGTCTCAGGCTCAGG 636
 QY 187 cggggcgccggcgagggcccccgggggtcccgggcggtctctggggactttctactc 246
 DB 635 TGGGGCGGCGGCTGGACGCGGACGGGTGGCACAGCTGGCATACGCGTCCCTCCACAGG 576
 QY 247 tacctggccaatctggaggccaagagcgaggtggcgcgctgctcctccccggggc 306
 DB 575 TGGCGGTAGACGCGCGCGGGACGCGGACCAACAGGGGGCGCCAGCAGCGCCAGCAGC 516
 QY 307 cgaagagtgcccaacgacgagctcttcggggcggtccagactcagcgacagtggtg 366
 DB 515 AGGCGGCGGCGCAGGGCGGGCTGGCGAGCGCGGCGCCAGGAGGGGGGTGACTCG 456
 QY 367 asgtggccatcttcagccacatgcacgacagtgctgcacgactcttccccggggca 426
 DB 455 AGGCAGCGCTCAGGCTGAGCAGGCGCGGTGAGCAGCAGCTGGCGTACATGCTGAGCGG 396
 QY 427 asgtactgtgacacatgtaccagctcaccagggcccccgccacacctcttg 480
 DB 395 CACAGTAGTACACCGCTTGACGCGCGCTGGCCCGCTGGCCAGCGCCAGGCTGCCGG 342

RESULT 13
 X53491/c
 ID X53491 standard; DNA; 114955 BP.
 XX
 AC X53491;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.
 XX
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 XX
 OS Synthetic.
 OS
 XX W09913886-A1.
 XX
 XX 25-MAR-1999.
 XX
 XX 17-SEP-1998; 98WO-US19419.
 XX
 XX 09-JUN-1998; 98US-0093972.
 XX 17-SEP-1997; 97US-0059160.
 XX
 XX (UYEC-) UNIV EAST CAROLINA.
 XX
 XX Nyce JW;
 XX
 XX WPI; 1999-229400/19.
 XX
 XX New antisense oligonucleotides used in treatment of, e.g. pulmonary
 XX vasoconstriction
 XX
 XX Disclosure; Page 37; 120pp; English.
 XX
 XX The specification describes antisense oligonucleotides (X52869-X55271)
 XX directed against at least 2 mRNAs selected from target genes, coding and
 XX non-coding regions of RNAs corresponding to target genes, gene
 XX initiation codons, genomic flanking regions, intron-exon borders, the
 XX 5'-end, the 3'-end and the juxta-section between coding and non-coding

KW MORT-1; Hf1; FAS/AP01 receptor; FAS-R; tumour; cancer; HIV;
KW mediator of receptor toxicity; gene therapy; ss.

OS	Homo sapiens.
XX	
PH	Key
FT	Location/Qualifiers
FT	1..771
FT	/*tag= a

[illegible]

Query Match 2.8%; Score 40; DB 17; Length 1701;
Best Local Similarity 49.1%; Pred. No. 4.9;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Search completed: December 7, 2000, 18:10:10
Job time: 26043 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 13:43:21 ; Search time 3361.48 Seconds
(without alignments)
1847.608 Million cell updates/sec

Title: US-09-434-708-3

Perfect score: 1422

Sequence: 1 atgctctgcgtggccccc.....ccaggaccctgcccgcc 1422

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pri.*

10: gb_pr2.*

11: gb_pr3.*

12: gb_ro.*

13: gb_sy.*

14: gb_un.*

15: em_fun.*

16: em_hum1.*

17: em_hum2.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: gb_htg1.*

31: gb_htg2.*

32: gb_in1.*

33: gb_in2.*

34: em_ba1.*

35: em_ba2.*

36: em_hum3.*

37: em_hum4.*

38: gb_pr4.*

39: gb_htg3.*

40: gb_htg4.*

41: gb_htg5.*

42: gb_htg6.*

43: gb_htg7.*

44: em_htg1.*
45: em_htg2.*
46: em_htg3.*
47: em_hum5.*
48: gb_pl3.*
49: gb_pr5.*
50: gb_htg8.*
51: gb_htg9.*
52: gb_htg10.*
53: gb_htg11.*
54: gb_htg12.*
55: gb_htg13.*
56: gb_htg14.*
57: gb_in3.*
58: gb_htg15.*
59: gb_htg16.*
60: gb_htg17.*
61: em_htg4.*
62: em_htg5.*
63: em_htg6.*
64: em_htg7.*
65: em_hum6.*
66: gb_htg18.*
67: gb_htg19.*
68: gb_htg20.*
69: gb_htg21.*
70: gb_htg22.*
71: gb_htg23.*
72: gb_vil.*
73: gb_vil2.*
74: gb_ba3.*
75: em_htg8.*
76: em_htg9.*
77: em_htg10.*
78: em_htg11.*
79: em_htg12.*
80: em_htg13.*
81: em_htg14.*
82: em_htg15.*
83: em_htg16.*
84: em_htg17.*
85: em_htg18.*
86: em_htg19.*
87: em_htg20.*
88: em_htg21.*
89: em_htg22.*
90: em_htg23.*
91: gb_pr6.*
92: gb_pr7.*
93: gb_sts1.*
94: gb_sts2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1418.8	99.8	1584	11	AF117646	AF117646 Homo sapi
2	1415.6	99.5	1575	9	AB028645	AB028645 Homo sapi
3	1132.8	79.7	1446	11	AF117647	AF117647 Homo sapi
4	340	23.9	3090	91	HSCCBL	X57110 Human mRNA
5	329.2	23.2	2661	57	DMU87925	U87925 Drosophila
6	327.6	23.0	2529	57	DMA3175	AJ23175 Drosophila
7	323.6	22.8	3241	91	HSU26711	U26711 Human cbl-b
8	323.6	22.8	3354	91	HSU26712	U26712 Human cbl-b
9	323.6	22.8	3982	91	HSU26710	X57111 Mouse mRNA
10	270.6	19.0	2808	12	MMCCBL	J04169 Mouse Cas N
11	248.2	17.5	2711	72	CASNS1	AC073760 Mus muscu
12	214.2	15.1	237653	60	AC073760	

```

13 199.4 14.0 2230 33 CEDNASL11
14 179.2 12.6 36890 41 AC014994
15 179.2 12.6 139226 39 AC010033
16 179.2 12.6 281936 32 AE003555
17 159.8 11.2 208239 51 AC021988
18 127.2 8.9 1758 4 AF237765
19 98 6.9 32559 33 CELM02A10
20 81.2 5.7 173038 70 AP001003
21 81.2 5.7 177595 70 AP001557
22 76.4 5.4 169523 41 AC015788
23 76.4 5.4 179680 11 AC016138
24 56.8 4.0 4855 2 SAU03114
25 51.8 3.6 11212 53 AC025047
26 49.4 3.5 117183 1 AE001986
27 49 3.4 63734 1 AF127374
28 48.6 3.4 87283 51 AC022856
29 47.8 3.4 40352 2 SCC61A
30 47.6 3.3 12801 2 SCF6
31 47.2 3.3 107444 66 AC078780
32 46.6 3.3 2510 73 SH1UL2XA
33 46.6 3.3 3797 73 PVULSGENE
34 46.6 3.3 9382 73 SH1ULGNS
35 46.4 3.3 16951 2 MSU46844
36 46.4 3.3 208918 60 AC074335
37 46.2 3.2 807 91 HSKCNQF01
38 45.6 3.2 34962 2 SCG20A
39 45.2 3.2 149402 67 AL157400
40 45 3.2 7913 3 BOVCWPR
41 44.8 3.2 15559 1 AF074603
42 44.8 3.2 40355 2 SCL6
43 44.8 3.2 169054 52 AC023767
44 44.6 3.1 9999 1 AE001942
45 44.2 3.1 10419 1 AE001865

```

ALIGNMENTS

```

RESULT 1
LOCUS AF117646 1584 bp mRNA PRI 17-JUN-1999
DEFINITION Homo sapiens long CBL-3 protein (CBL-3) mRNA, alternatively spliced
isoform, complete cds.
ACCESSION AF117646
VERSION AF117646.1 GI:4959420
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1584)
AUTHORS Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M.,
Penninger,J. and Lipkowitz,S.
TITLE Cbl-3: a new mammalian cbl family protein
JOURNAL Oncogene 18 (22), 3365-3375 (1999),
MEDLINE 99289203
REFERENCE 2 (bases 1 to 1584)
AUTHORS Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M.,
Penninger,J. and Lipkowitz,S.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-1999) Genetics Dept, Medicine Branch, DCS, NCI,
Bethesda Naval Hospital, Bethesda, MD 20889, USA
FEATURES
source
1..1584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2"
/cell_line="CF-PAC1"
/tissue_type="pancreatic adenocarcinoma"
1..1584
/gene="CBL-3"
64..1488
CDS

```

```

/gene="CBL-3"
/note="long form, alternatively spliced; cbl family
protein; structurally related to the sli-1 protein,
encoded by the cbl gene of Caenorhabditis elegans"
/codon_start=1
/product="long CBL-3 protein"
/protein_id="RAD34341.1"
/db_xref="GI:4959421"
/translation="MALAVPWQWEERARALGRAVRLQRLEEQVDPRLSVSPSL
RDLIPRAQLLRVARRAGGGPGGPGGDFLLIYLANLEAKSRQVAALLPPRG
RRSANDLEFRAGSLRLQAKLAIIFSHMAELHALFPYGYGHMYQLKAPAFIFW
RSCGACRVLPWAEFSLGTHFVEPCGTALARTIDLTCSGHYSIFEDVETRLF
QWPTLLKNWOLLAVNHPYMAFTYDEVQERLQACRDKPGSIFRSPCLGOWAIG
YVSDSGSLQITIPANKPLSOVLLEGOKDGVLYPDGKTHNPDLTELGOAEPQRIHVS
BEQOLYWMDSHFELCKIAESKNKYKIEPCGHLHLCSCCLAAWQSDSOTCFRCRE
IKGWEAVSITQFGATAEQSDSSDQEGRELELQGVPLSAPPLPRPDUPPRKPRNA
QPKVRLKNGSPPAALGPDPAPA"
BASE COUNT 303 t 538 c 482 g 261 t
ORIGIN
Query Match 99.8%; Score 1418.8; DB 11; Length 1584;
Best Local Similarity 99.9%; Pred. No. 5.1e-242;
Matches 1420; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atggctctggcgggtggccccgtggggcgacagtgggaagagggccccgcctggggcgg 60
Db 64 ATGGCTCTGGCGGTGGCCCCGTGGGGCGGACAGTGGGAAGAGGCGCGCCCTGGCGCG 123
QY 61 gaagtcaagagctgcagcgcctagaagagcaatcgctgacccccgcctgtcctgaagt 120
Db 124 GCAGTCAGGATGCTGCAGCGCTAGAGAGCAATGCTCGACCCCGCGCTGCTCGTGAGT 183
QY 121 cccctctgcctggggacctctgccccgcacagcagctgcttcagaggtggcccat 180
Db 184 CCCCCTTCGCTGGGGACCTGCTGCCCGGACACGCGAGCTGCTTCAGAGGTGGCCCAT 243
QY 181 tctggcgggcgccggcggaggcgccccgggggtcccgggcggtctctggggactttcta 240
Db 244 TCTCGCGCGGCGCGCGGAGGCGGCGCGCGCGGCGGCGCTCTGGGGACTTTCTA 303
QY 241 ctcatctacctggcdaatctgagggcccaagagagcagatggccgctgcctccccc 300
Db 304 CTCATCTACCTGGCCCAATCTGGAGGCGCAAGAGAGGAGGTGGCGCGCTGCTGCTTCCC 363
QY 301 cggggcggaagagatgccaacgacgagctcttcggggcggtctccagactcagcgacag 360
Db 364 CGGGCGCGAAGAGTGGCCAAAGCAGAGCTCTTCGGGGCGGGTCCAGACTCAGCGGACAG 423
QY 361 ctggccaagctggccatcatcttcagccacatgcacacagatgcagcactctcccc 420
Db 424 CTGGCCAAGCTGGCCATCATCTTCAGCCCATGCACGAGAGTGCAGGACTTCTTCCCC 483
QY 421 ggggcaagtactgtgacacatgtaccagctaccagctcccaagggccccgcacacctctgg 480
Db 484 GGGGGAAGTACTGTGGACACATGTACCAGCTACCAAGGCCCCCGGCCACACTTCTGG 543
QY 481 agggaaagtctcgagacccgggtgtgtctccctgggtgagttgagttccctccctgggc 540
Db 544 AGGAAAAGTTGCGAGGCGCGGTGTGTCTGCCCTGAGTTTGAAGTCCCTCCCTGGGC 603
QY 541 acctggcacacctgtggaaacaggctgcacagcctggccttgcgcaccacctgacctc 600
Db 604 ACCTGCCACCTGTGGAACACGAGCTGCACAGCCTGGCCTTGGCCACCACTTGGCTC 663
QY 601 acctgcagcggcagctgtgctcatcttcagttgagttgagtttccaccaggtctttacgcor 660
Db 664 ACCTGCAGCGGCGCAGTGTGCATCTTCAGTTCGAGTCTTCCACCAAGGCTCTTTCAGCCA 723
QY 661 tggccaacacctcccaagaactggcagctcctctgagttgagttgagttgagttgagttgagtt 720
Db 724 TGGCCACACCTCTCAAGAACTGGCAGCTCCTGGCAGTCAACCAACCAAGCTACATGGCC 783

```

```

Qy 721 ttcctacattgatgaggtccaaagaggtctgaaggctcagggacagacagccaggcagt 780
Db 784 TCCCTCACCTATGATGAGTCCAAAGAGCGTCTCAGGCGTGCAGGGAACAAGCCAGGCAGT 843
Qy 781 tacattcttcggccagctgtactgcctcgggagcagtgaggccatcggtctatgtgagctca 840
Db 844 TACATCTTCGGCCAGCTGTACTCGCTGGGAGTGGGCCATCGGTATGTGAGCTCA 903
Qy 841 gatggagcatctgcagaccatctcctgcacaaacacccctctccagggtgctcctggag 900
Db 904 GATGGAGCATCTCGAGACCATCCCTGCCAACAACCCCTGCCAGGTGCTCCTGGAG 963
Qy 901 gacagaagagcggtcttcacctctaccagatggaagacccacacacacacacacacac 960
Db 964 GGCACAAGGACGGCTTCTACCTCTACCCAGATGGAAGACCCACACACACACACACAC 1023
Qy 961 gagctcggcaggcagacacccacagcagcgcacatccacgtctcagagagcagctgcagctc 1020
Db 1024 GAGCTGGCAGCAGACACCCAGCAGCAGCATCCAGTGTCTAGAGGAGCAGCTGCAGCTC 1083
Qy 1021 tactggccatggactccacatttgactctgcaagatctgtcgtgagggcaacaagat 1080
Db 1084 TACTGGCCATGACTCCACATTTGACCTCTGCAAGATCTGCTGAGAGCAACAGGAT 1143
Qy 1081 gtgaagattgagcgtcggggccacctgctgcagctgctgctggtcctggcagc 1140
Db 1144 GTCAAGATTGAGCGCTGCGGACCTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203
Qy 1141 tcgacagccagacctgcctctcctcctcctcctcctcctcctcctcctcctcctcctcct 1200
Db 1204 TCGGACAGCAGACCTGCCCCCTTCTCCGCTGCGAGATCAAGGCGCTGGAGGCGCTGAGT 1263
Qy 1201 attaccagttccacggtcaggtctactgtcgtgagactcagggaacagcagtgaccaggaa 1260
Db 1264 ATCTACAGTTCTACGGTCTAGGCTACTGCTGAGGACTCAGGGAACAGCAGTACACAGGAA 1323
Qy 1261 ggacggagttgagctggggcagggtgcccctcttgcctcctcctcctcctcctcctcctc 1320
Db 1324 GGCAGGAGTTGAGCTGGGCGAGTGGCCCTTTGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1383
Qy 1321 gactcctcccccaggagccagaaatgccagccagaaagttagactcctaaaggagaa 1380
Db 1384 GATCTGCCCCCAGGAGGCCAGAAATGCCAGCCGAAAGTAGACTCTTAAGGGGGAAC 1443
Qy 1381 tccctccagctgctgggaccccgagccctgcccctgcccggcc 1422
Db 1444 TCCCTCCAGCTGGCTGGGACCCAGGACCTTGGCCCGGCC 1485

RESULT 2
AB028645 1575 bp mRNA PRI 10-NOV-1999
LOCUS Homo sapiens mRNA for Cbl-c, complete cds.
DEFINITION AB028645
ACCESSION AB028645
VERSION AB028645.1 GI:6291531
KEYWORDS Cbl-c.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kim,M., Tezuka,T., Suzuki,Y., Sugano,S., Hirai,M. and Yamamoto,T.
TITLE Molecular cloning and characterization of a novel cbl-family gene,
Cbl-c
JOURNAL Gene 239, 145-154 (1999)
REFERENCE 2 (bases 1 to 1575)
AUTHORS Kim,M., Tezuka,T. and Yamamoto,T.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1999) to the DDBJ/EMBL/GenBank databases, Tadashi
Yamamoto, Institute of Medical Science, University of Tokyo,
Department of Oncology, Shirokanedai 4-6-1, Minato-ku, Tokyo
108-8639, Japan (E-mail:tyamamoto@ims.u-tokyo.ac.jp,
Tel:++81-3-5449-5301, Fax:++81-3-5449-5413)

```

Location/Qualifiers

1..1575

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/map="19q13.2-3"

42..1466

/gene="cbl-c"

42..1466

/gene="cbl-c"

/codon_start=1

/product="Cbl-c"

/protein_id="BAA86298.1"

/db_xref="GI:6291532"

/translation="MALAVAPGWQREEARALGRAVRLQRLLEECQVDPRLSYSPESL
RDLLPRTAQLLREVAHRAAGGGPGGPGDGLLIIVLANEAKSRQVALLPRLG
RRSANDLFRAGSELRLQAKLAIIEFHMHAELHALFPGKYCGHMYQLTKAPAHFW
RESCGARCLVPLAEFESLGTCHPVEPGCTALALRTILDTSGSHVSIIEFVFTLE
QPMPTLLKNQLLAVTHPGTMAFLTYDEVQERLQACDRKPGSIIFPSCTIRGQWALG
YVSDGSILOTIPANKPLSOVLGQDGFYIYPDKGTNPDLTELGQAEPQRIHVS
EEQLIYWAMDSITFELCKIAESNKDKIIEPCGHLKSCCLAAQWSDSQTCPFCRCE
'IKWEAVSYVQFYQATAEQNSNSDQEGRELELGVPLSAPPLPRPDLPRKPRNA
QPKVRLKNSPPAALGPDPAPA"

BASE COUNT 316 a 527 c 474 g 258 t

ORIGIN

Query Match 99.5%; Score 1415.6; DB 9; Length 1575;

Best Local Similarity 99.7%; Pred. No. 1.9e-241;

Matches 1418; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 atggtcttggcggg-ggccccctggggggcagctgggaagagcccgccctggggcgg 60

Db 42 ATGGCTCTGGCGGTGGCCCCCTGGGGCGACAGTGGGAAGAGCGCCGCCCTGGCGCG 101

Qy 61 cagctcagatgctgcagcgcctagaagagcaatgctcgacccccggctcctcgtagt 120

Db 102 GAGTCAAGATGCTGCAGCGCTAGAAAGCAATGCGTCGACCCCGGCTGTCGTGAGT 161

Qy 121 cccctctcgtcggggacctgctccccgcacagcagctgcttcagagagtggtggccat 180

Db 162 CCCCCTTCGCTGGCGGACCTGTCGCCCGCACAGCGAGCTGCTTCGAGAGTGGCCAT 221

Qy 181 tctggcggggggccggggagggggccggggggtcccgcgactctgggagatttcta 240

Db 222 TCTCGCGGGGCGCGCGCGGCGCGCGCGCGGGGTCCCGCGGCTCTGGGGAGCTTCTA 281

Qy 241 etcatctacctggccaatctggaggccaagagcagcaggtggccggcgtcgtgcctccc 300

Db 282 CTCATCTACCTGGCCAACTTGGAGGCCAAGAGCAGGAGGTGGCGGCTGCTGCTCCC 341

Qy 301 gggggccaaaggagtgcacaagcagagctcttcggggggggtccagactcaggcgacag 360

Db 342 CGGGGCCGAAAGAGTGCACAGCAGAGCTCTCCGGCGGGCTCCAGACTCAGCGGACAG 401

Qy 361 ctggccaaagctggccatcatcttcagccacatgcaacgagcagcagcactcttcccc 420

Db 402 CTGGCCAGCTGGCCATCATCTTCAGCCACATGACGAGAGCTGCAGCGACTCTTCCC 461

Qy 421 gggggcaaaagtactgtggacacatgtaccagctaccacaggcccccccgccacacatttgg 480

Db 462 GGGGAAAAGTACTGTGGACATGTACAGGTACCAAGAGGCCCGCCGACACCTTCTGG 521

Qy 481 agggaaagtgcggagccgggtgtgtgcctcgggtggtgagtttgagctcctcctgggc 540

Db 522 AGGGAAGTTCGGAGCGCGGTGTGTGTCCTCGCTGGCTTGAGTTTGATCTCCTGGGC 581

Qy 541 acctgccacctgtggaaccagctgcacagccctggccttgcgcaccaccattgacctc 600

Db 582 ACCTGCCACCTGTGGACACAGGCTGCACAGCCCTGGCTTGGCACCACCATTCACCTC 641

Qy 601 acctgcagcggcagcgtgtccatcttcagtttcagcttcaccacagcgtcttttcaagca 660


```

VERSION      U26711.1  GI:862408
KEYWORDS
SOURCE
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 3241)
AUTHORS      Keane.M.M., Rivero-Lezcano.O.M., Mitchell.J.A., Robbins.K.C. and
               Lipkowitz.S.
TITLE        Cloning and characterization of cbl-b: a SH3 binding protein with
               homology to the c-cbl proto-oncogene
JOURNAL      Oncogene 10 (12), 2367-2377 (1995)
MEDLINE      95303504
REFERENCE    2 (bases 1 to 3241)
AUTHORS      Lipkowitz.S., Keane.M.M. and Mitchell.J.A.
TITLE        Direct Submission
JOURNAL      Submitted (10-MAY-1995) Stan Lipkowitz, Navy Medical Oncology
               Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval
               Hospital, Bethesda, MD 20889, USA
FEATURES
   source
       1..3241
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /cell_line="breast cancer cell line ZR75-1"
           /map="3q"
           /chromosome="3"
       323..2755
           /note="binds SH3 proteins; similar to c-cbl proto-oncogene
               product, Swiss-Prot Accession Number P22681"
           /codon_start=1
           /product="cbl-b truncated form 1"
           /protein_id="AAB09292.1"
           /db_xref="GI:862409"
           /translation="MANSNGRNPGRGGRKGRILGIIIDATQDVGPPKQAAARRR
               TVEKWLMDKVLQCNPKLQKNSPPYILDLPYQHLRLILSLSDNOKLAQLS
               ENYKFIIDSLMKKRAIRLFKEGKERMEYEQSDRRNLKLSIFSHMLAEIKAI
               FPNQFQDNFRITKADAAEFKRFKFDKTVPKVFRCLHEVHVISLDMALKS
               TIDLNDYISVFEEDFTLRFQWGSILRNWFLAVHPGYMAFTYDEVKARLQKY
               STRGSGYIFRLSLTGQWALGYVTGDNILQIPENKPLFQALIDGSGEGLYPDG
               RSYNPDLTGCEPTPHDIKVTQEQYELICEMGSTFQCKICAKENKDKVILPFCGLM
               CISLQAWQSDGCGPCFCEKNGTEPIIVDFDPEDEGRCCSIDFEGFMDLDD
               DDDRESLMNLNANVKCTDRNSPVTSPGSPSLAQRKQPDPDLQIFHLSPVP
               PRDLIQGVIRSPGSGPTGSPKSPKVMKQKPLPAPPPLRDPPLPPPPPPPIIP
               PNRLSHIIHVSIVSPRDPMPLEAWCRDVFNTQNLVGLRLLGSGSPKPTASNN
               VNGHSVSDPYLMRKRRLDPLLEGAKVFSNHLGSEYDVPRLSPPPVPTLLP
               SIKTGPLANSLEKTRDPVEDDDDEYKIPSSHPVSLNSOPSHCHNVKPVVSCDNGH
               CMLNTHGPSESKSNIPDLISYLKGVDFDSADPVLPPARPPTRDNPKHSSLNRT
               PSDLDLILPPLG"
       668..679
           /note="encodes nuclear localization signal"
       1439..1558
           /note="encodes ring finger"
       2751
           /note="alternative splice at nucleotide 2751 truncates
               protein before leucine zipper"
BASE COUNT   907 a 758 c 720 g 856 t
ORIGIN
Query Match      22.8%; Score 323.6; DB 91; Length 3241;
Best Local Similarity 59.48; Pred No. 3.4e-48;
Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;
Qy 351 caggcgacagctggcagctgcatcatcttcagccacatgcacgcagagctgcacgc 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 739 CAGACGAATCTCACAACCTGTCCTATCTTCAGTCACATGCTGCAGAAATCAAGC 798
Qy 411 actcttccccgggcaagtagtcttgacacatgtaccagctaccagagccccgccca 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db* 799 AATCTTTCCCAATGGTCAATTCAGGAGATAACTTTTCGTATCACAACACGATGTCG 858
Qy 471 cacctctcggaggaaagttcggagcccggtgtgtctgcctgggctgagttcagtc 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 859 TGAATTCGGAGAAAGTTTTTTTGGAGACAAAACATATCGTACCATGGAAGTATTGAGACA 918
Qy 531 cctcctggggcaccctgcaccctgtggaccagctgcacagccctggccttgcgcaccac 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 919 GTCCCTTCATAGGTCACACAGATTAGCTCTAGCTGGAGCAATGGCTCTAAATCAAC 978
Qy 591 cattgacctcucctgcagcgggcaogtgcctatcttcctgagttgcagctcttccaccagct 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 979 AATTGATTAACTTGAATGCAATGATTACATTTTCAGTTTGAATTTGATTTTACCAGGCT 1038
Qy 651 ctctcagccatggcacaactcctcaagactggcagctcctcctggcagtcacacccacag 710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1039 GTTTCAGCCTTGGGGCTCTAATTTGGGGAATTGGAAATTTCTTAGCTGTGACACATCCAGG 1098
Qy 711 ctacatggcctctcctcaccctatgatgaggtccaaagagtgctgcagggctgcagggacaa 770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1099 TTACATGGCATTTCTCACATATGATGAAGTTAAAGCAGACACTACAGAAATATAGACCAA 1158
Qy 771 gccagcagttacatcttccggccacagctgtactgcctggggcagtggtggccatcggcta 830
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1159 ACCCGGAAGCTATATTTTCCGGTTAAGTTGCACCTCGATTGGGACAGTGGGCCATTTGGGCTA 1218
Qy 831 tctgagctcagatggcagcatcctcgcagacatcctcctgcacacacccctgtcccaggt 890
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1219 TGTGACTGGGGATGGGAATATCTTACAGACCATACTCATACAGCCCTTATTTCAAGC 1278
Qy 891 gctcctgggggacagagagcaggtcttacctctaccacagatggaagaccacacaccc 950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1279 CCTGATTGATGGCAGCAGGGAAGGATTTTATCTTTATCCTGATGGGAGAGTTATAATCC 1338
Qy 951 aaacctgactgagctggccagcaggaacccacagcagcagcagcagcagcagcagcagcag 1010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1339 TGATTTAATGATGATTATGTGAACCTACACCTCATACCATATAAAGTTACACAGGAACA 1398
Qy 1011 gctgcagctcctactggggccatggaactccacatttgactctgcaagatctgtgtgagag 1070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1399 ATATGAATTAATTTGTAATGGGCTCCACTTTTCAGCTCTGTAAGATTTGTGCAGAGAA 1458
Qy 1071 caacaagatctgaagattgagccctggcggcagcagcagcagcagcagcagcagcagcag 1130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1459 TGACAAAGATGTCAGATTGAGCCTTGTGGGCATTGTGACCTTTCCTTACGGC 1518
Qy 1131 ctggcagcactcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1519 ATGGCAGGAGTCGGATGGTCAGGGCTGCCCTTTCGTGCTGTGAAATAAAGGAACTGA 1578
Qy 1191 ggcctgagctatctaccagttccacaggtccacaggtcagcagcagcagcagcagcagcag 1250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1579 GCCCAATATCGTGGGCCCTTTGATCCAAAGATGAAGGCTCCAGGTTGTTCAGCATCAT 1638
Qy 1251 tgaccaggaagcagggaggttg 1272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1639 TGACCCCTTGGCATGCCGATG 1660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT      8
HSU26712    3354 bp      mRNA      PRI      30-SEP-1996
LOCUS      Human cbl-b truncated form 2 lacking leucine zipper mRNA, complete
DEFINITION cbl-b
ACCESSION  U26712
VERSION    U26712.1  GI:862410
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3354)
AUTHORS    Keane.M.M., Rivero-Lezcano.O.M., Mitchell.J.A., Robbins.K.C. and
            Lipkowitz.S.
TITLE      Cloning and characterization of cbl-b: a SH3 binding protein with
            homology to the c-cbl proto-oncogene
JOURNAL    Oncogene 10 (12), 2367-2377 (1995)

```

MEDLINE 95303504
 REFERENCE 2 (bases 1 to 3354)
 AUTHORS Lipkowitz, S., Keane, M.M. and Mitchell, J.A.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-1995) Stan Lipkowitz, Navy Medical Oncology Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval Hospital, Bethesda, MD 20889, USA
 FEATURES Location/Qualifiers
 source 1..3354
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="breast cancer cell line ZR75-1"
 /map="3q"
 /chromosome="3"
 CDS 323..2635
 /note="similar to c-b1 proto-oncogene product, Swiss-Prot Accession Number P22881"
 /codon_start=1
 /product="cbl-b truncated form 2"
 /protein_id="AAB09293.1"
 /db_xref="GI:862411"
 /translation="VANSNGENPGGRGNPKRGILGILDAIDAVGPPKQAAADR
 TVETWKLMDKVRICQPKLQKNSPPYILDLDTYQHLRLILSKYDDNQLAQLS
 ENEYKIIYDLSMKSKRAIRLFKGRKMEYEQSDRRNLTUKLSLISHMLAEIKAI
 FPGQFQDNFRTIKADAAEFWRKFGDKTIIVPKVFRQCLHEVHQISSLEAWALKS
 TDLACNDYISVEFIEDIRLFPQPGSILRNWFLAVHPGYMAFLTYDEVKARLOKY
 STPKGSYIFSLCTELGOWAIGYVTDGNILOTIPHNKPLFQALIDGREGFYLYPDG
 RSYNPLDGLCPTPDHDKVTOEYELVCEMGSTFOLCKICAENDKVKIPECGHLM
 CTSCLTAQESDQGCPCRCFIKTEPIIIVDPDPDESGSCSIIIDFGMWLIDL
 DDDREESLMNRLANVKCDTRQNSPVISPGSSPLAARRPQDPLPIRLHSLPPVP
 PRDLTKQIVASPCGSPSPKSVKVRKODKPLPAPPPPLRPDPPLPPPPPPPIIP
 PNRLSRHHVSEPPDPMPLEAWCPDRVFTNQLVGRLLGEGSPKPIGTASSN
 VNGRHSRVGSDSLMRKRRHDLPLEGAKVFNHGLSEEDYDVPRLSPPPPVITLLP
 STKCTGFLANSLEKTRDPVEDDDEYKIPSSHPVSLNSQPSHCNRPVPRSDNGH
 CMLNTHGSPSEKSNIPDLSYILKGYRI"
 misc_feature 668..679
 /note="encodes nuclear localization signal"
 misc_feature 1439..2023
 /note="encodes ring finger"
 misc_feature 2619
 /note="alternative splice at nucleotide 2619 truncates protein before leucine zipper"
 BASE COUNT 943 a 773 c 739 g 899 t
 ORIGIN
 Query Match 22.8%; Score 323.6; DB 91; Length 3354;
 Best Local Similarity 59.4%; Pred. No. 3.3e-48;
 Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;
 Qy 351 caggcgacagctggcccaagctggccatcatcttcagccacatgcacgcagagctgcacgc 410
 Db 739 CAGACGAAATCTCACAAAACGTCCCTTATCTTCAGTCACATGCTGCAGAAATCAAGC 798
 Qy 411 actcttccccgggcaaaagtactgtgacacatgtaccagctcaccaaggccccgccca 470
 Db 799 AATCTTCCCAATGGTCAATCCAGGAGATACCTTCGTATCACAAGACGATGCTGC 858
 Qy 471 cactcttgaggagaaagtgtcgagagcccggtgtgtgctgccttggcctggagttgagtc 530
 Db 859 TGAATCTGGAGAAAGTGTTCGAGACAAAACATACTGACCATGGAAGATATTCAGACA 918
 Qy 531 cctctctgggacactggccaccttggaaaccagctgcacagccctggccttggccaccac 590
 Db 919 GTGCGCTTCATGAGTCACACAGATTAGCTCTAGCGTGAAGCAATGCTCTAAAATCAAC 978
 Qy 591 cattgacctcacctggcaggggcagctgtccatctctcagttcagcttcacccaggt 650
 Db 979 AATGATTTACITGCAATATACATTTTCAGTTTGTGATTTGATTTTACCAAGCT 1038
 Qy 651 ctttcagccatggccacacactcctcaagaactggcagctcctctggcagtcacaccaggt 710
 Db 1039 GTTTCAGCCTTGGGGCTCTATTTTTCGGGAATTTGAAATTTCTTACGTGTGACATCCAGG 1098

Qy 711 ctacatggccttcctcactatgatgaggtcccaagagcgtctcagcgcttcgagggacaa 770
 Db 1099 TTATGATGCAATTTTTCATATGATGAAGTTAAAGCAGCACTACAGAAATATACCAAC 1158
 Qy 771 gccaggagcttacatcttcctggccacagctgtactcgcctggggcagtgggccatcgcta 830
 Db 1159 ACCCGAAGCTATATTTTCGGTTAAGTTGACATCGATGGGACAGTGGCCATTGGCTA 1218
 Qy 831 tctgagctcagatggcagcagcatcctgcagagaccatccctgcacacacacccctgtccaggt 890
 Db 1219 TGTGACTGGGATGGGATATCTTACAGACCATACCTCAATAACAAGCCCTTATTTCAGC 1278
 Qy 891 gctcctggaggacagaaaggcgtctcactctaccagatggaaagaccccaacccc 950
 Db 1279 CTTGATTGATGGCAGCAGGGAAGGATTTTATCTTTATCTTGATGGGAGGATTATAATCC 1338
 Qy 951 agactgactgagctggccaggccagaaaccccccagcagcaccatcgtctcagaggagca 1010
 Db 1339 TGATTTAATCTGATTTGTAACCTACACCTCATGACATATAAAAGTTACACAGGAACA 1398
 Qy 1011 gctcagctctactggccagctgactccacattgagctctgcaagatcgtgctgagag 1070
 Db 1399 ATATGAATATTTGTGAATGGGCTCCACITTTTCAGCTCTGTAAGATTGTGCAGAGAA 1458
 Qy 1071 caacaaggatgtgaagattgagccgtggcgacactgctcgcagctgctgcctggctgc 1130
 Db 1459 TGACAAAGATGTCAGATTTGAGCCTTGTGGCATTTGATGTGCACCTCTTGCCTTACGGC 1518
 Qy 1131 ctggcagcactgcagacagcagcagcagcctcctctgcctgcagatcagagctggga 1190
 Db 1519 ATGGCAGAGTCGGATGTCAGGGCTGCCCTTTCTGCTGTGAAATAAAGGAATGA 1578
 Qy 1191 ggcctgagttatccaccgttccacggtccagctactgctgagcagcaggaacagcag 1250
 Db 1579 GCCCATATCTGGACCCCTTTGATCCAGAGATGAAGGCTCCAGGTTTTCAGCATCAT 1638
 Qy 1251 gacccagaaaggccaggaggtg 1272
 Db 1639 TGACCCCTTGGCATGCCGATG 1660
 RESULT 9
 HSU26710 3982 bp mRNA PRI 30-SEP-1996
 LOCUS Human cbl-b mRNA, complete cds.
 DEFINITION U26710
 ACCESSION U26710
 VERSION U26710.1 GI:862406
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3982)
 AUTHORS Keane, M.M., Rivero-Lezcano, O.M., Mitchell, J.A., Robbins, K.C. and Lipkowitz, S.
 TITLE Cloning and characterization of cbl-b: a SH3 binding protein with homology to the c-b1 proto-oncogene
 JOURNAL Oncogene 10: (12), 2367-2377 (1995)
 MEDLINE 95303504
 REFERENCE 2 (bases 1 to 3982)
 AUTHORS Lipkowitz, S., Keane, M.M. and Mitchell, J.A.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-1995) Stan Lipkowitz, Navy Medical Oncology Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval Hospital, Bethesda, MD 20889, USA
 FEATURES Location/Qualifiers
 source 1..3982
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="breast cancer cell line ZR75-1"
 /map="3q"
 /chromosome="3"

```

CDS      323..3271
/note="interacts with SH3 proteins; similar to c-cbl
proto-oncogene product, Swiss-Prot Accession Number
P22681"
/codon_start=1
/product="cbl-b"
/protein_id="AAB09291.1"
/db_xref="GI:862407"
/translation="MANSNGRNPGCGGNPKRGLIGLIDAIQDAVGPPOAAADRR
TVETKWLMDKVVLCQNPKLQKLNPPKILDPDTPYQHLRLILSKYDDNOKLAQLS
ENFYKIIYDLSMKRKRRLKEGKERMEYEQSDRRNLTKLSIFSHMLAEIKAI
FPGQFGDNDFRTKADAAEFWRKFQDKTIVWKFVROCLHEVHQISSLSDAMALKS
TIDLTQNDYISVEFFIDFTLQFQWGLNWNFLAVTHPGVMAFTYDEVKARLQKY
STKPGSVIFRLSCTRILGQWALGVYTGDNILQITPHKPLFOALIDGSEGVLYPDG
RSYNPDLTGLCESTPHDHLKVTOEQYELICEMGSTFQCKICAENDKDKVIRPCEHLM
CTSLTANQSDGCGCFKCEKLEPIIVDFPRDGSRCSSLIIDFGMPMDLDD
DDDRSESLMNNLANVKCTDRQNSPVSPPGSSPLAQRKPPDPQLPHLSLPVP
PRDLIQGIVRPGSPGSPKSPCMVKQDKPLPAPPPLRDPFPPPPRPPPIIP
PNRLSRHHIVHVESVPSRPPMPLEAWCPDRVFGTQNLVGCRLLEGSSPKPGITASSN
VNGRHGRVSDPVLMDKRRHRLPLEGAKVSNGLHSGSEYDVVPELSPKPPVPTLLP
SIKCTGLANLSLSEKTRDPEVEDDEYKIPSSHPVSLNSQPSHCNKPVPSCDNGH
CMLNGTHGPSSEKSNIPOLSIYLLKGDVEDSDPVLPPARPPTRDNDPKXHSLSNRT
PQYDILLPLPGDADALPPLSPPPPPFARHSLIEHSPGSSSRPSGQDLFLPS
DPVDLASQGVPLPPARLPGENVKNRTSQDYDQLPSCDSQAPARPPKPRRTA
PEIHRKPHGPEALENDVDAKIAMGEGAFEEVKRALEIAQNNVVEARSILREFAF
PPVPSPLNL"
misc_feature 668..679
/note="encodes nuclear localization signal"
misc_feature 1439..1558
/note="encodes ring finger"
misc_feature 3116..3223
/note="encodes leucine zipper"
BASE COUNT 1093 a 969 c 877 g 1043 t
ORIGIN

Query Match      22.8%; Score 323.6; DB 91; Length 3982;
Best Local Similarity 59.4%; Pred. No. 3.1e-48;
Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 351 caggcgacagtgcccaagctggccatctcattcagccacatccacgacgagctgcacgc 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 739 CAGACGAAATCTCAAAATGTCCTTATCTTCAGTCACATCTGTCGAGAAATCAAGC 798

QY 411 actttcccccgggcaagtactgtggacacatgtaccagctcaccaggcccccacca 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 799 AATCTTTCCCAATGGTCAATTCAGGAGAGATAACTTTCGTATCACAAGACAGATGCTG 858

QY 471 cactctcagagaaagtgcgagcccggtgtgtgctgcctggctggctgagttgagtc 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 859 TGAATTCGAGAAAGTTTTTTGGAGACAAACTATCGTACCATGGAAGATATTCAGACA 918

QY 531 cctcctgggacactgcacacccctgtggaaccaggctgcacagccctggcttgcgaccac 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 919 GTGCTTTCATGAGTCCACAGATTAGCTCTAGCCTGGAAAGCAATGGCTCTAAATCAAC 978

QY 591 cattgaacctacactgcaggggacgctgtccatcttcgagttcgagctgtttcaccaggct 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 979 AATTGATTTAACTTGAATGATTACATTTTCAAGTTTTTGAATTTGATTTTACCAGGCT 1038

QY 651 ctttcagccatggccacacactcctcaagaactggcgagctcctggcagttcaaccaccagg 710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1039 GTTTCAGCTTGGGGCTCTATTTCGGAATTTGGAATTTCTTAGCTGTACACATCCAGG 1098

QY 711 ctacatggccttctcactcatatgataggtccaaagcgctctgcagccctgcagggacaa 770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1099 TTACATGGCAATTTCTCACATATGATGAAGTTAAAGCACCACATACAGAAATATAGCACC 1158

QY 771 gccaggcgagttacattctccggcccgagctgtactccctcggggcgagtcggcctcggcta 830
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1159 ACCGGAGAGCTATATTTCCGGTTAGTTGCACTTCGATTTGGGACAGTGGGCCATGGCTA 1218

QY 831- tgtgagtcagatgagcagcatcctgcagaccatccctgcacaaacccctgtcccaggt 890

```

Db	1011	TGTTACTGCCGATGGGARACATTTCTGCAGACAAATCCACACAATAACCGCTCTTCCCAAGC	1070
QY	891	gctcctggaggagacagaagagcggttctacctctaccagatggaagaccacaccc	950
Db	1071	ACTGATTGATGGCTTCAGGGAAGGCTTCTATTGTTTCTCTGAGAGAAATCAAAATCC	1130
QY	951	agacactgactgaactcgccagggcagaaccccgagcgatccacgtgtcagaggagca	1010
Db	1131	TGACCTGACAGGTTTATGTGAACCACTCCCAAGATCATATCAAAATGAC	1181
QY	1011	gctgcagctctactggcgcatggactccacatttgagctctgcaagatctgtctgagag	1070
Db	1182	-----CCAGATATGCTGTGAGAA	1199
QY	1071	caacaagatgtgaagattgagcgtgogggcaactctctgagctgctcctggtgtgc	1130
Db	1200	TGATGAAGATGTGAAGATTGACCCCTGGACACTCATGTGCACATCCTCCCTCAGCTC	1259
QY	1131	ctggcagcactgcgacagccagacctgccccttgcgcgtcgagatcaaggctggga	1190
Db	1260	GTGGCAGGAATCAGAAGTCAAGGCTGCTCCCTTTTGGGATGTGAAATCAAGGTA	1319
QY	1191	ggcctgagttatctaccagttccac	1215
Db	1320	GCCCATCTGTGTGATCGCGTTTGAC	1344
RESULT	11		
CASINS1			
LOCUS			
DEFINITION	Mouse Cas NS-1 retrovirus gag-onc fusion protein (v-cbl) gene, 3'		
ACCESSION	J04169		
VERSION	J04169.1		
KEYWORDS	gag-onc fusion protein; viral oncogene.		
SOURCE	CAS NS-1 retrovirus DNA, isolated from mouse (strain NFS/N) previously infected with Cas-Br-M virus.		
ORGANISM	Cas NS-1 murine leukemia virus		
REFERENCE	Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses; 1-Mammalian type C virus group.		
AUTHORS	1 (bases: 1 to 2711)		
TITLE	Langdon, W.Y., Hartley, J.W., Klinken, S.P., Russetti, S.K. and Morse, H.C.III.		
JOURNAL	v-cbl, an oncogene from a dual-recombinant murine retrovirus that induces early B-lineage lymphomas		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 86, 1168-1172 (1989)		
COMMENT	89445204		
FEATURES			
source	1. ..2711		
	/organism="Cas NS-1 murine leukemia virus"		
	/db_xref="taxon:11793"		
CDS	41..1609		
	/note="gag-onc fusion protein"		
	/codon_start=2		
	/protein_id="AAA42885.1"		
	/db_xref="GI:323270"		
	/translation="LERLKLSQVHPDPEDPFGGTNVKSF IWQSPADIGRKLRLERL		
	DLKSLTGDLVREAKIFNKRETPREERIKRETKEREERADREQRERREHREH		
	REMSKLATVVGQFQGGERRERPOLDDQCAVCKETHGWAASAGGRRGFSFSP		
	SGPSLAERADDPPLAWAGNVKSSGSGGSGGAGGLGLMKDAFOPHHHHHL		
	SHPPCTVDKMKVKWKLMDKVLQCNPNVALKNPPYILDLDPYIOHLRTVLSRYE		
	YEGKMTIGENEYFVFMENLMKTKQIISLFKEGKERYEENSQPRNLTKLISF		
	HMLAKLGIFSGFGQGTFTRTKADAAEFWRKAFGEKTIIPWKSFQALHEVHPIS		

GLEAMALKSTIDTCDNDYSVFEEDITRLFPQWSSLLRNNSLAVTHPGYMAFLYD
 EVARLQKFTHKQSGVIFRLSCIRLQGAWGYVTADGNLIQTIPHNKPLEQALIDGFR
 EGYLPFDGRNQNPDLTGCEPHTS"
 752 a 688 c 688 g 583 t
 BASE COUNT
 ORIGIN 1 bp upstream of XbaI site.

BASE COUNT	701:a	487 c	429 g	613 t
ORIGIN				
Query Match	14.0%;	Score 199.4;	DB 33;	Length 2230;
Best Local Similarity	53.5%;	Pred. No. 3.8e-26;		
Matches 466;	Conservative 0;	Mismatches 396;	Indels 9;	Gaps 2;
QY	353	gycacagctggccaagctggccatcatcttcagccacatgacgcagagctgcacgcaac	412	
DB	531	GACGGAAGCTTACGAAATGCTACTACATTTTCACATATGCTTTTCGAGATTAAGCAT	590	
QY	413	tcttcocggggcaaatctgtggacacatgacacagctcaccagggccccgcgccaca	472	
DB	591	TATTTCCGGAAGGATATCTATTGAAACCGGTTTCGGATGACAAAGGAGCGGANA	650	
QY	473	ccctctggaggaaaagttggagccgggtgtgctgccctgggctggagtttgatgcc	532	
DB	651	GCCTTTTGAGTCATCAITTTTACAAAAAACAATGTACCTGGTCACATTTTTTACGT	710	
QY	533	tectgggcaactgcca--ccctgtgaaccaggtctcacagccctggccttgcgcaca	589	
DB	711	CATTAGAAAGCACCATGGATCAACGATAGGAAAAATGGAAGCAGCGGAATTAAGAGCTA	770	
QY	590	ccattgacctcactgcagcgcgctgccatcttcagttgcagctcttcaccagcg	649	
DB	771	CGATAGACTTGCGCGGAGATGATTTTATTCGAATTTTGAGTTTGATGTGTTTACAAGGT	830	
QY	650	tctttcagccatggccacacactcctcaagaactggcagctcctggcagtcacaccaccag	709	
DB	831	TATTCACCTTTCAAAACACGTGATCAAAATTTGGCAACACTCACACCGCCCATCCCG	890	
QY	710	gctacatggctctcactcatgatgaggtccaaagcgctgcagggcctgcagggaca	769	
DB	891	GATATGTGCATTTCTCATATACATGAGTCAAAAACGGTTAGAAAAATTAACGANA	950	
QY	770	agccaggcagttacatcttcggcccgccagctgactgcctggggcgagtgggccatcggt	829	
DB	951	AACTCTGGAAGCTACATCTTCGGTTATATGACACACTCTGTGACAAATGGCAATAGGAT	1010	
QY	830	agtgagctcagatggcagcactcctgcagaccatccctgcacaacacccctgtcccag	889	
DB	1011	ACGTAGCTCCGATGGAAAGATTTATCAGACATATACCACAGATAAAGTTTGATTCAG	1070	
QY	890	tgctcctggaggacagaagcagcgtctacctctaccagatggaaagacccacacc	949	
DB	1071	CACATCATGAAGGCCATAAAGAAGATTTATATTTACCGGAACGGTAGAGATCAAGATA	1130	
QY	950	cagacctgac:gagctcggcagcagacacccacagcgcatccagtgctcagaggagc	1009	
DB	1131	TTAACTTATCCAAATGATGATGGCATGGCCACAGCGGCACAGTGCAGTGCACATGAC	1190	
QY	1010	agctgcagctctactggccgcagctccacatttgagctctcagaagattctgtcgtgaga	1069	
DB	1191	AATACGAGTTGTATTTGTGAGATGGCACACATTCGAGTTGTGCAAAATTTGTGACGATA	1250	
QY	1070	gcacaacaggatgtgaagattgagccgtgcgggcaactgctcagactgctcctgctg	1129	
DB	1251	ACGAGAAGAACATCAAAATGAGCCATGTGGACATTTGCTCTCGCAAAATGTTTGCTTA	1310	

QY	1130	ctgtggcagcaactcga-----cagccagacactccctcttctgcgcgtgcagatcaagg	1183
Db	1311	ACTGGCAGGATTCGGATGGTGGTGGCAACATCTCCATCTTCGGGTACGAAATCAAG	1370
QY	1184	gctgggagggccgtgagtattaccagttcca	1214
Db	1371	GACAAATCGTGTGATTATTCACAGGTTCAA	1401
RESULT 14			
AC014994/c			
LOCUS	AC014994	36890 bp	DNA
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.	HTG	16-NOV-1999
ACCESSION	AC014994		
VERSION	AC014994.1	GI:6436341	
KEYWORDS	HTG; HTGS_PHASE2.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 36890)		
TITLE	Adams,M. and Venter,J.C.		
JOURNAL	Direct Submission		
COMMENT	Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA This sequence was identified as CDM:10212090 by the submitter. For further information on this sequence e-mail to fly@celera.com. * NONE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.		
FEATURES	Location/Qualifiers		
source	1..36890		
	/organism="Drosophila melanogaster"		
BASE COUNT	11572 a 7373 c 7329 g 10616 t		
ORIGIN			
Query Match	12.8%;	Score 179.2;	DB 41; Length 36890;
Best Local Similarity	65.5%;	Pred. No. 5.3e-23;	
Matches 262;	Conservative	0; Mismatches 138;	Indels 0; Gaps 0;
QY	506	tgtgtgcctggcgtgagtttgagtcctctctgtggcagctgccacctgtgtggaaccagct	565
Db	30388	TGGTTCCTCGAAATCTTCGGCAGGAGGTTTACAAGTACATCCCATATCTCCGGCC	30329
QY	566	gcacagccctggccttcgcgcacaccattgacctcaactgcagggcgagtgctccatct	625
Db	30328	TGGAGGCCATGGCCCTAAAGACCACATTCGATCTTACCTGCACGAGCTTCTATTTCCAACT	30269
QY	626	tcagtttcgacgtctttcaacaggctctttcagccatcggccaaactcctcaagaactggc	685
Db	30268	TCGAGTTCGACGCTTCTTCACAGGCTCTTCACAGCTTCGTCAGCTTCGGGTACCCCTCAGCAACTGGC	30209
QY	686	agctctggcagtcacacccacccaggctacatggcctctcaccattgatgaggtccaaag	745
Db	30208	AGATTCTGGCCGTACACATCCGGGCTACGTGGCGTTTCTACATACGACGAGGTGAAGG	30149
QY	746	agcgtctgcaggcgtgcagggaacaagccagcagtgattacatctctccggccagctgtactc	805
Db	30148	CTCGCCTACAGCGCTACATCCTCAAGCGGGCAGCTACGTTTTCCGGGCTCTCTCTGCACGC	30089
QY	806	gcctggggcgagtgggccattcggtatgtgagctcagatggcagctactcgcagaccatcc	865
Db	30088	GATTGGGCCAATGGCCATCGCTACGTAACTCCGAGGGAGAGATTCTCGACACAATCC	30029
QY	866	ctgccacaacacccctgtccaggtgctcctctgaggagaca	905
Db	30028	CTCAGAACAGTGCCTGTGTCCGAGCGCGCTGTCTCATGGCCA	29989

13397 13496: gap of unknown length
13497 17696: contig of 4200 bp in length
17697 1796: gap of unknown length
1798 21860: contig of 4064 bp in length
1799 21860: contig of 4064 bp in length
21861 21960: gap of unknown length
21962 25098: contig of 3138 bp in length
25099 25198: gap of unknown length
25199 27960: contig of 2762 bp in length
27961 28060: gap of unknown length
28062 30796: contig of 2736 bp in length
30797 30896: gap of unknown length
30898 33631: contig of 2735 bp in length
33632 33731: gap of unknown length
33732 36353: contig of 2622 bp in length
36354 36453: gap of unknown length
36454 38018: contig of 2565 bp in length
38019 39118: gap of unknown length
39119 41507: contig of 2389 bp in length
41508 41607: gap of unknown length
41609 43946: contig of 2339 bp in length
43947 44046: gap of unknown length
44048 46295: contig of 2249 bp in length
46296 48395: gap of unknown length
48396 48993: contig of 2198 bp in length
48994 48993: gap of unknown length
48995 50813: contig of 2120 bp in length
50814 50913: gap of unknown length
50914 53017: contig of 2104 bp in length
53018 53117: gap of unknown length
53118 53095: contig of 1978 bp in length
53096 55195: gap of unknown length
55196 57155: contig of 1960 bp in length
57156 57255: gap of unknown length
57256 59209: contig of 1954 bp in length
59210 59309: gap of unknown length
59310 61213: contig of 1904 bp in length
61214 63133: gap of unknown length
63134 63167: contig of 1854 bp in length
63168 63267: gap of unknown length
63268 65097: contig of 1830 bp in length
65098 65197: gap of unknown length
65198 66930: contig of 1733 bp in length
66931 67030: gap of unknown length
67031 68684: contig of 1654 bp in length
68685 68784: gap of unknown length
68786 70393: contig of 1609 bp in length
70394 70493: gap of unknown length
70494 72081: contig of 1588 bp in length
72082 72181: gap of unknown length
72182 73749: contig of 1568 bp in length
73750 73849: gap of unknown length
73850 75387: contig of 1538 bp in length
75388 75487: gap of unknown length
75488 77000: contig of 1513 bp in length
77001 77100: gap of unknown length
77101 78600: contig of 1500 bp in length
78601 78700: gap of unknown length
78701 80199: contig of 1499 bp in length
80200 80299: gap of unknown length
80300 81739: contig of 1440 bp in length
81740 81839: gap of unknown length
81840 83268: contig of 1429 bp in length
83269 83368: gap of unknown length
83369 84792: contig of 1424 bp in length
84793 84992: gap of unknown length
84994 86250: contig of 1358 bp in length
86251 86350: gap of unknown length
86351 87676: contig of 1326 bp in length
87677 87776: gap of unknown length
87778 89094: contig of 1318 bp in length
89095 89194: gap of unknown length
89195 90449: contig of 1253 bp in length
90450 90549: gap of unknown length

90550 91801: contig of 1252 bp in length
91802 91901: gap of unknown length
91903 93127: contig of 1226 bp in length
93128 93227: gap of unknown length
93228 94440: contig of 1213 bp in length
94441 94540: gap of unknown length
94541 95730: contig of 1190 bp in length
95731 95830: gap of unknown length
95831 97015: contig of 1185 bp in length
97016 97115: gap of unknown length
97116 98210: contig of 1095 bp in length
98211 98310: gap of unknown length
98311 99370: contig of 1060 bp in length
99371 99470: gap of unknown length
99471 100529: contig of 1059 bp in length
100530 100629: gap of unknown length
100630 101680: contig of 1051 bp in length
101681 101780: gap of unknown length
101781 102825: contig of 1045 bp in length
102826 102925: gap of unknown length
102926 103948: contig of 1023 bp in length
103949 104048: gap of unknown length
104049 105054: contig of 1006 bp in length
105055 105154: gap of unknown length
105155 106153: contig of 999 bp in length
106154 106253: gap of unknown length
106254 107244: contig of 991 bp in length
107245 107344: gap of unknown length
107345 108328: contig of 984 bp in length
108329 108428: gap of unknown length
108429 109408: gap of unknown length
109409 109508: gap of unknown length
109509 110479: contig of 971 bp in length
110480 110579: gap of unknown length
110580 111536: contig of 957 bp in length
111537 111636: gap of unknown length
111637 112557: contig of 921 bp in length
112558 112657: gap of unknown length
112659 113571: contig of 914 bp in length
113572 113671: gap of unknown length
113672 114559: contig of 888 bp in length
114560 114659: gap of unknown length
114659 115542: contig of 883 bp in length
115543 115642: gap of unknown length
115643 116516: gap of unknown length
116517 117461: contig of 845 bp in length
117462 117561: gap of unknown length
117562 118402: contig of 841 bp in length
118403 118502: gap of unknown length
118503 119341: contig of 839 bp in length
119342 119441: gap of unknown length
119442 120279: contig of 838 bp in length
120280 120379: gap of unknown length
120380 121210: contig of 831 bp in length
121211 121310: gap of unknown length
121311 122141: contig of 831 bp in length
122141 122141: contig of 831 bp in length

Query Match 12.8%; Score 179.2; DB 39; Length 129226;
Best Local Similarity 65.5%; Pred. No. 3.4e-23;
Matches 262; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 506 tgctgcctggctgagttgagtcctctctgggacactgccacctgtggaaaccaggct 565
Db 122325 TGGTTCCTG3AAAATCTTCGGCAGGAGGTAAACAAGTACATCCCATATATCCGGCC 122384
Qy 566 gcacagccctggcctgcgcacacacattgacctcactgcagcggcagcgtgtccatct 625
Db 122385 TGGAGGCCATGGCCCTAAAGACCACACTATCGATCTTACCTGCACAGGACTTCAATTCACAACT 122444
Qy 626 tcgagttcgagttcttcacacaggtctttcagcctcctggcgcacacactcctcagaaactggc 685
Db 122445 TCGAGTTCGAGTCTTCACAGCCTCTTCACAGCCTTGGGTGACCTGCTAGGCACAACTGGC 122504

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 06:02:31 ; Search time 79.91 seconds
(without alignments)
2977.917 Million cell updates/sec

Title: US-09-434-708-1
Perfect score: 1547
Sequence: 1 cggaggtccatggtctt.....caaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 73620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	340	22.0	3090	6	PCT-US93-06251-7
2	46.4	3.0	9960	5	US-08-822-586-46
3	42.2	2.7	1693	5	US-09-320-878-23
4	41.6	2.7	3292	5	US-09-320-878-22
5	40	2.6	1701	5	US-09-357-072-1
6	40	2.6	1701	6	PCT-US95-16342-1
7	40	2.6	1701	6	PCT-US96-10521-1
8	38.4	2.5	4692	3	US-08-916-917-1
9	38.4	2.5	4692	3	US-08-972-631-1
10	38.4	2.5	4692	3	US-08-972-639-1
11	38.4	2.5	4692	3	US-08-972-630-1
12	38.4	2.5	4692	3	US-08-672-211-1
13	38.4	2.5	4692	3	US-09-225-170-1
14	38.2	2.5	954	5	US-08-680-506-5
15	38.2	2.5	1347	5	US-08-680-506-8
16	38.2	2.5	1587	5	US-08-680-506-6
17	38.2	2.5	2233	5	US-08-680-506-4
18	38.2	2.5	5027	5	US-08-680-506-2
19	38	2.5	2150	5	US-09-263-083-1
20	38	2.5	5392	3	US-08-403-852D-1
21	38	2.5	5392	5	US-08-510-645B-1
22	38	2.5	15378	5	US-08-785-420-1
23	38	2.5	44377	3	US-08-804-237C-7
24	38	2.5	44377	4	US-08-804-198-1
25	37.4	2.4	53526	5	US-08-658-136-2
26	37.4	2.4	53577	5	US-08-658-136-1

27 37.2 2.4 3596 3 US-08-779-801-5 Sequence 5, Appli
28 37.2 2.4 3632 3 US-08-779-801-3 Sequence 3, Appli
29 37.2 2.4 3632 3 US-08-779-801-4 Sequence 4, Appli
30 36.8 2.4 1174 5 US-09-034-985-1 Sequence 1, Appli
31 36.8 2.4 3293 4 US-08-442-899A-75 Sequence 75, Appli
32 36.6 2.4 4451 5 US-08-717-294-42 Sequence 42, Appli
33 36.4 2.4 9723 2 US-08-083-590A-21 Sequence 21, Appli
34 36.4 2.4 9723 5 US-08-532-384-21 Sequence 21, Appli
35 36 2.3 1312 1 US-08-305-506A-1 Sequence 1, Appli
36 36 2.3 1312 6 PCT-US94-02389-1 Sequence 1, Appli
37 36 2.3 2558 2 US-08-477-476-1 Sequence 1, Appli
38 36 2.3 2558 4 US-08-993-210-1 Sequence 1, Appli
39 36 2.3 2558 6 PCT-US96-07354-1 Sequence 1, Appli
40 36 2.3 152331 5 US-09-128-155-16 Sequence 16, Appli
41 35.8 2.3 2286 5 US-09-176-657-4 Sequence 4, Appli
42 35.8 2.3 5996 5 US-08-528-934-33 Sequence 33, Appli
43 35.6 2.3 1560 4 US-08-709-959A-2 Sequence 2, Appli
44 35.6 2.3 2030 4 US-08-530-950-1 Sequence 1, Appli
45 35.6 2.3 2277 2 US-08-676-967-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1
PCT-US93-06251-7
; Sequence 7, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3090 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-06251-7

Query Match 22.0%; Score 340; DB 6; Length 3090;

Best Local Similarity 62.2%; Pred. No. 1.8e-69;

Matches 533; Conservative 0; Mismatches 325; Indels 0; Gaps 0;

QY 364 agggagcagctggccagctggccatctttcagccacatgcacgagctgcaqca 423

DB 590 AGGGAAACCTAACCAACTGTCCTCATCTTCAGCCACATGCTGGCAGAACTAAAGGA 649

FILE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 23
LENGTH: 1693
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-23

Query Match 2.7%; Score 42.2; DB 5; Length 1693;
Best Local Similarity 48.9%; Pred. No. 0.37; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 118

QY 809 gctgtactgcgtgggcagtgccatcgctgtgtgagctcagatgacagcatctgc 868
DB 971 GCTGGAGGTCCAGGTTCGGGTACGGGTACGGCAAGCGCGAGCCGCTTCGTCG 912
QY 869 agacatccctcccaaacacccctgtctccaggttctctctgagggagacagcgtc 928
DB 911 AGACCTTCGGCCACCAAGATCGAGACCGAGAAATCGAGAACCCATCGCGAGGTCT 852
QY 929 ttatctctcaccagatgaaagacccacacccagacctgactgagctcggccagcag 988
DB 851 TCGACCTCCGCCCGCGCGGATCATCGCGACCTCGACCTGCTCGCGGATCTACTCC 792
QY 989 aacccagcagcgtccatccatgctgtagagagcagctgagctgagctgagcgtc 1039
DB 791 AGACCGCGCTACGGCCACTTCGGCGCGAGCTCGCGGACTTCACCTGGG 741

RESULT 4
US-09-320-878-23/c
Sequence 22, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 3292
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-22

Query Match 2.7%; Score 41.6; DB 5; Length 3292;
Best Local Similarity 46.3%; Pred. No. 0.61;
Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 85 ctgcagcgcctagaaagacaatcgctgcaccccccggcgtctccgtgagtcctcccttgcgtg 144
DB 3248 CCGCTCGGCCCATTCGCGCACGCGTGCATCACCCGAGCGCTCGCGGCGCTCCAGGTG 3189
QY 145 cgggacctgctgccccgcacagcgacgtctcttcgagaggtggcccatcttcggcgggcg 204
DB 3188 CGGCGCATCGGACGGCTGAGGACCTGCGCGGGAAGCTCTCGCGCCCGCGGAGCGAGCC 3129
QY 205 gcggcgagggcgccccgggggtcccgggggtctctgggaacttttactatctactcctg 264
DB 3128 TTCGGCGGTTCCTCGCCCGCTAGCGGCGGAGAGGTGCACGGGTAGTGGCT 3069
QY 265 gccaatctggagccaaagacgagcgaggtggcgcgctgctgctcccgggggccgaag 324
DB 3068 GAGGTCTCATCCCGCGGCGCTCGAGTGCCTCGCGAGCTCGTCGGGCGCTCGGTGCG 3009
QY 325 agtccaaacagcagctcttcccggggggtccagactcagcgacagctggccaa 380
DB 3008 CACGGTGAAGAGGTGCCAGACGCGGTGCGTGTGGCGCGGTACCGGCGAGCGCGA 2953

RESULT 5
US-09-357-072-1
Sequence 1, Application US/09357072
Patent No. 6015712
GENERAL INFORMATION:
APPLICANT: Brett P. Monia
APPLICANT: Brenda F. Baker
APPLICANT: Hong Zhang
APPLICANT: Lex M. Cowser
TITLE OF INVENTION: ANTISENSE MODULATION OF FADD EXPRESSION
FILE REFERENCE: R1S-0027
CURRENT APPLICATION NUMBER: US/09/357,072
CURRENT FILING DATE: 1999-07-19
NUMBER OF SEQ ID NOS: 87
SEQ ID NO 1
LENGTH: 1701
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (145)..(771)
US-09-357-072-1

Query Match 2.6%; Score 40; DB 5; Length 1701;
Best Local Similarity 49.1%; Pred. No. 1.2;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 44 agtggaaagagcccgccctggcgcccgagtcagatgctgcagcgcctagaagagc 103
DB 242 agcgaagctgagcgcgtgacgagcgcttagacctcttccatgctgctgagcaga 301
QY 104 aatcgctgcaccccgctgctccgtgagtcctcccttcgctgcgggacctgctgcccgcga 163
DB 302 acgacctggaagcccgccacacccagatctctgcgagctgctgcctccctcgcgcc 361
QY 164 cagcgagctgcttcgagaggtggcccatcttcggcgggcgccggcgagggcgcccg 223
DB 362 acgacctgctcgcgcgctgacgacttcgagggggggggcgccggcgccgctg 421

Qy 224 ggggtccgaggcgctctgggaactttctactcatct 259
||| | | | | |||
Db 422 qqaqaaagacatctgtgcagcatttaacgtcatat 457

```

RESULT      6
PCT-US95-16542-1
; Sequence 1, Application PC/TUS9516542
; GENERAL INFORMATION:
; APPLICANT: VEDA RESEARCH AND DEVELOPMENT CO. LTD.
; APPLICANT: WEINWURZEL, Henry
; APPLICANT: WALLACH, David
; APPLICANT: BOLDIN, Mark
; APPLICANT: VARFOLOOMEV, Eugene
; APPLICANT: METT, Igor
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APOL1
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16542

```

```

Query Match      2.6%: Score 40; DB 6; Length 1701;
Best Local Similarity 49.1%: Pred. No. 1.2;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 44 agtgggaagagcccgccgacctggccggcgagtcgagcgacctagaagagc 103
DB 242 AGCGCAAGCTGAGCGCGTGCAGAGCGGCTGACCTCTTCTCCATGCTGCTGGAGCAGA 301
QY 104 aatgcgctgacccccggcgctcccgtagtcgccctctcgctggcgagcctgctgccccca 163

```

Db	302	ACGAGCTGGAGCCCGGCACACCGAGCTCCTCGCGAGCTGCTGCCTGCCCTGCGGGCC	361
Qy	164	csagcagctgctcgagagtggtgccattctcggcgagcgccgcgagcgagcgccccg	223
Db	362	ACGAGCTGCTGCGGGCGGTGCGACATTCGAGCGGGGGCGGCGCGCGCGCCTG	421
Qy	224	ggsgtccccggggtotggggactttctactcatct	259
Db	422	GGGAGAGAGACTGTGTCCAGCACTTAAAGCTCAT	457

```

RESULT 7
PCT-US96-10521-1
; Sequence 1, Application PC/TUS9610521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..768
PCT-US96-10521-1

```

	Query Match	2.6%	Score 40;	DB 6;	Length 1701;
	Best Local Similarity	49.1%;	Pred. No. 1.2;		
	Matches 106;	Conservative 0;	Mismatches 110;	Indels 0;	Gaps
QY	44	aftgggaagagigcccgcgcctcggccggcgagtcagagatctcagcgctctagaagac	103		
Db	242	AGCGCAAGCTGAGCGCGGTGACAGCGGGCTAGACCTCTTCGATGCTCTGGACGAGA	301		
QY	104	aatgcgtctgacccccggcgtgcgtgagtcctcccttcgctcgggacctctctcccgcga	163		
Db	302	ACGACCTGAGCCCGGGCACACCGAGACTCTCTCGCGAGACTGCTGCCTCCCTCGCGGCGC	361		
QY	164	cagcgcgagctctcttcgagaggtggccattcttcggcgggcgccggcggagggcgcccg	223		
Db	362	ACGACCTGCTTCGCGCGGTCTGACACTTCGAGCGGGGGCGCGCGCGCGCGCGCTTC	421		

QY 224 ggggtccgggggtggtggaactttctactcatct 259
|||||
Db 422 GGAAGAAGACCTGTGTGACGACATTTACGTCATAT 457
|||||
RESULT 8
US-08-916-917-1/c
; Sequence 1, Application US/08916917
; Patent No. 5856132
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip Thomas
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,917
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/672,211
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8549-0006-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-916-917-1

Query Match 2.5%; Score 38.4; DB 3; Length 4692;
Best Local Similarity 50.9%; Pred. No. 3.6;
Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
QY 16 gctctggcggtggcccgctgggggacagtggaagagccgagcccgccctggcgcgga 75
|||||
Db 1833 GCTTGGGCTGGGACGACGAGCGGGAG-CCTGGGCCCGGGACGGGGTCGGGTG 1775
|||||
QY 76 gtacgatgctgcagcgccctagagagcaatgcgtgcaccccggtgtccgtgagtc 135
|||||
Db 1774 TCCAGGGGCTGCAGAGGCTGCCGGCCCGACGAGGGCGGGCTGCCGGTGC 1715
|||||
QY 136 ccttcgctgcgggacgtgtgccccacacagcgacgtcttcgagagtggtgcccattct 195
|||||
Db 1714 TCCAGGCTGCAGAGTCCCTCCGACACCATCTGGCTGTGCTCTTGAAGATTATAG 1655
|||||
QY 196 cggcgggcgggcgagagcgcccggggtcccgcggtctctggg 243
|||||
Db 1654 ATCCTGTGAATCTCTGCCCGGGCGGGTGGCCCGGGCTCGTG 1607
|||||

RESULT 10
US-08-916-917-1/c
; Sequence 1, Application US/08916917
; Patent No. 5856132
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len

RESULT 9
US-08-972-631-1/c
; Sequence 1, Application US/08972631
; Patent No. 5856133
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip T.
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,631
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/672,211
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 8549-0005-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)854-3660
; TELEFAX: (415)854-3694
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-972-631-1

Query Match 2.5%; Score 38.4; DB 3; Length 4692;
Best Local Similarity 50.9%; Pred. No. 3.6;
Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

QY 16 gctctggcggtggcccgctgggggacagtggaagagccgagcccgccctggcgcgga 75
|||||
Db 1833 GCTTGGGCTGGGACGACGAGCGGGAG-CCTGGGCCCGGGACGGGGTCGGGTG 1775
|||||
QY 76 gtacgatgctgcagcgccctagagagcaatgcgtgcaccccggtgtccgtgagtc 135
|||||
Db 1774 TCCAGGGGCTGCAGAGGCTGCCGGCCCGACGAGGGCGGGCTGCCGGTGC 1715
|||||
QY 136 ccttcgctgcgggacgtgtgccccacacagcgacgtcttcgagagtggtgcccattct 195
|||||
Db 1714 TCCAGGCTGCAGAGTCCCTCCGACACCATCTGGCTGTGCTCTTGAAGATTATAG 1655
|||||
QY 196 cggcgggcgggcgagagcgcccggggtcccgcggtctctggg 243
|||||
Db 1654 ATCCTGTGAATCTCTGCCCGGGCGGGTGGCCCGGGCTCGTG 1607
|||||

RESULT 10
US-08-972-629-1/c
; Sequence 1, Application US/08972629
; Patent No. 5859201
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len


```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,211
FILING DATE: 27-JUN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-672-211-1

```

Query Match	2.58;	Score 38.4;	DB 3;	Length 4692;
Best Local Similarity	50.94;	Pred. No. 3.6;		
Matches 116;	Conservative 0;	Mismatches 111;	Indels 1;	Gaps 1;
QY	16	gtctgcggtgccccctggggcgacagtggaaagagcccgccctggccggcca	75	
Db	1833	GTCTGGCTGGGGCAGCAGCGGGAG-CCCTGGCCCGGACGGGGCAGGTCGGGCTG	1775	
QY	76	gtcaggatgctgcagcgccctagaagcaatgctgaccccccggctgcctgtagtccc	135	
Db	1774	TCCAGGGGCTGCAGAGGCTCCGCGCCGACGACAGGGCGGCGCTGCCGGCTGCC	1715	
QY	136	ctttgcgtgcggacctctgcccgacacagcgagctgttcgagagtgggccattct	195	
Db	1714	TCAGGCTCGCGAGTCCTCCGACAGACCATCTGGTGTGTCCTTGAAGATTATAG	1655	
QY	196	cggcgggggcgccggagggcccccggggtcccgcgggctctggg	243	
Db	1654	ATCCTGTTGAATCTGTCGGCGGCGCGGGTGGCCCGGCGGCTCGTGG	1607	

```

RESULT 13
US-09-225-170-1/c
; Sequence 1, Application us/09225170
; Patent No. 6017763
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip Thomas
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/225,170
; FILING DATE:
; CLASSIFICATION:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,917
FILING DATE: 15-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Adams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCES/DOCKET NUMBER: 8549-0006-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4632 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-225-170-1

Query Match : 2.5%; Score 38.4; DB 5; Length 4692;
Best Local Similarity 50.9%; Pred. NO.3.6;
Matches 116; Conservative 0; Mismatches 111; Indels 1;
Gaps 1;

Qy	16	gctctgcgctgccccctgsgggcgacagtggaagagccccgcgccttggccggcgca	75
Db	1833	GCTTGGCTGGGCGACGACGCGGAG-CCTTGGGCCCGGAGCGGGGCGAGGTCGGCTG	1775
Qy	76	gtcagatctcagcgccctagaagaacatgctcacccccgcctgtccgtgagtcctc	135
Db	1774	TCCAGGGGGCTGCAGAGCTGCCGCCGACGACAGGGCGGGCGCTGTCCGGGCTGCC	1715
Qy	136	ccttcgctcgggacctgctgcccgcacagcagctgctctcgagaggtggcccatct	195
Db	1714	TCACAGCTCGCGAGTCCCTCCGCGACCATCTGGCTGTGCTCTGAAGAGTTTATAG	1655
Qy	196	cggcggcgccggcggagcgccccggggggtcccggcgctctggg	243
Db	1654	ATCTCTGTAACTTGTCCCGGGCGCGGGTGGGCCCGGGCGCTCGTG	1607

RESULT 14
US-08-680-506-5/c

```

; sequence 5, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentId Ver. 2.0
; SEQ ID NO 5
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Gallus:gallus
; US-08-680-506-5

```

Query Match 2.5%; Score 38.2; DB 5; Length 954;
Best Local Similarity 50.2%; Pred. No. 2.6;
Matches 120; Conservative 0; Mismatches 118; Indels 1; Gaps 1;

Qy	115	ccccggctgtctcctgagtcctcccccttcctctgcgggaactgctgccccgcacacgcgcagctg	174
Db	390	CGCCAGGTGATCCATGTCGGGGGGGGCGCGCGCAGACCGCGCTCGAGGCTCTCCCTC	331
Qy	175	cttcgagagtggtggccattctcggcggggcgccggcgggagggcgccccgggggtcccgcc	234
Db	330	GTACCATCATCGCCAGACGCGGGCGGGCGGGCGAGGACGGGGCGCGCGCGCGG	271

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 10:56:07 ; Search time 152.48 Seconds
(without alignments)
3811.321 Million cell updates/sec

Title: us-09-434-708-1
Perfect score: 1547
Sequence: 1 cggaggctccctgctct.....caaaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 950044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	22.0	3090	20 X89079	Nucleotide sequenc
2	48	3.1	10732	21 A10594	Gene encoding a su
3	46.8	3.0	11955	20 X53491	Human adenosine Al
4	46.4	3.0	9960	19 V58939	Mycobacterium smeg
C 5	43	2.8	114955	20 X53491	Human adenosine Al
C 6	42.2	2.7	1693	21 X56005	Contig 004 from co
C 7	41.6	2.7	1140	21 X287290	S. venezuelae deso
C 8	41.6	2.7	3291	21 X56004	Contig 003 from co
9	41.6	2.7	12441	21 X287284	S. venezuelae deso
10	41.6	2.7	15613	21 X287319	S. venezuelae deso
11	41.4	2.7	1582	20 X08910	Human FADD protein
12	41.4	2.7	1642	17 T39397	FADD (Fas-associat

13	41.4	2.7	2644	20 X81941	CDNA encoding a pr
C 14	41.2	2.7	1170	20 Z39042	Human fishboy G-pr
C 15	40.4	2.6	880	17 T38913	B2-bradykinin rece
16	40.4	2.6	6225	20 X55273	Human enzyme-relat
17	40.4	2.6	6225	21 A34721	Human adenosine re
18	40	2.6	1701	17 T30372	MORT-1 cDNA. Homo
19	40	2.6	1701	18 T61397	MORT-1 coding sequ
20	40	2.6	1701	21 Z44745	Human FADD DNA. H
21	40	2.6	53789	19 V21187	Amycolatopsis medi
22	39.8	2.6	627	20 V19130	MORT1 isoform MORT
C 23	39.4	2.5	1308	17 T31602	Cartilage-derived
24	39.2	2.5	2170	21 X28402	Porcine betal-adre
25	39.2	2.5	15672	12 Q10613	Rianodin receptor
C 26	38.8	2.5	738	20 X86968	Ala acetacetyl-Co
27	38.8	2.5	3632	19 V61586	Alpha-1A calcium c
C 28	38.8	2.5	6436	20 X36985	DNA containing PHA
C 29	38.8	2.5	77536	21 A14651	Nucleotide sequenc
30	38.4	2.5	1746	11 Q05538	Sequence encoding
31	38.4	2.5	2914	20 X57059	Human KNO3 cDNA.
C 32	38.4	2.5	4692	19 V04633	Porcine phospholip
C 33	38.4	2.5	4692	20 V74099	Porcine G-protein
C 34	38.4	2.5	4692	21 Z86812	Pig p101 regulator
C 35	38.2	2.5	954	19 V19862	Gallus domesticus
C 36	38.2	2.5	1347	19 V19864	Gallus domesticus
C 37	38.2	2.5	1587	19 V19863	Gallus domesticus
C 38	38.2	2.5	2233	19 V19861	Gallus domesticus
C 39	38.2	2.5	5027	19 V19860	Sequence encoding
C 40	38	2.5	1999	11 Q04784	Mouse N-acetylgluc
C 41	38	2.5	2150	20 X87820	Tn5099-10 transpon
42	38	2.5	3318	15 Q88194	Sequence comprisin
C 43	38	2.5	5392	15 Q84201	MH mutant porcine
C 44	38	2.5	15377	13 Q25975	Platenolide synth
45	38	2.5	44377	18 T78508	

ALIGNMENTS

RESULT 1

X89079
ID X89079 standard; DNA; 3090 BP.
XX
XX X89079;
XX
XX 14-SEP-1999 (first entry)
DT Nucleotide sequence of human Cbl.
DE
KW LAT; tyrosine kinase; linker for activation of T cell; TCR; human;
KW T-cell receptor; TCR signalling pathway; neoplasia; inflammation;
KW hypersensitivity; allergy; microbial infection; genetic disease;
KW autoimmune disease; graft rejection; modulator; Cbl; ss.
XX
XX Homo sapiens.
XX
XX WO9932627-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-US27400.
XX
XX 23-DEC-1997; 97US-0068690.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Samelson LE, Zhang W;
PI Contig 004 from co
XX
XX WPI; 1999-418926/35.
DR P-PSDB; Y27126.
XX
XX Linker for activation of T cell protein used to, e.g. screen for
PT modulators of T cell signalling
XX

Db 104238 bggcbbggcgccgcccgcggcgccgcsnnndnnccgcbggcbbggcgccgcccgcgc 104297
 QY 182 aggtggccattctcggcgccgcccggcgagggcgcccccggggtcccgccggtctctg 241
 Db 104298 cgggcsnnndnnccgcbggcgccgcccgcggcgccgcsnnndnnccgcbggcbb 104357
 QY 242 gggactttctactatctacgtgcccattctggagggccacagagcagcaggtggccgcgc 301
 Db 104358 gggcgccgcccgcggcsnnndnnccgcbggcgccgcccgcggcgccgcsnnndn 104417
 QY 302 tgcctcctcccgcggcgccgaaagtgcacacgacgagcttctccggcggtccacagac 361
 Db 104418 nccgcbggcgccgcccgcggcgcsnnndnnccgcbggcgccgcccgcgcgc 104477
 QY 362 tcaggcgacgtggcgaagctggccatctcttcagccacatgcacgacagctgacg 421
 Db 104478 csnnndnnccgcbggcgccgcccgcggcgcsnnndnnccgcbggcgccgcccgcgc 104537
 QY 422 ca---ctctcccggggcgaagtaactgtgacacatgtaccagctcaccaagcccgcgc 478
 Db 104538 ccgcsnnndnnccgcbggcgccgcccgcggcgcsnnndnnccgcbggcgccgcccgcgc 104597
 QY 479 ccacacattctgaggggaaagtgcggagcccggtgtgtgctgcctggcggtgagttg 538
 Db 104598 gccsnnndnnccgcbggcgccgcccgcggcgcsnnndnnccgcbggcgccgcccgcgc 104657
 QY 539 agtcctctctggcactgcacccctgtggaac-cagctgcacagccctggccttcgcgc 597
 Db 104658 ndnnccgcbggcgccgcccgcggcsnnndnnccgcbggcgccgcccgcggcsnnndnnccgc 104717
 QY 598 accaccattgacctcactgcagcggcgacgtgtccatcttcgagttcgcagcttcacc 657
 Db 104718 bggcbbggcgcsnnndnnccgcbggcgccgcccgcggcsnnndnnccgcbggcgccgcsn 104777
 QY 658 aggtctttcagcactggcgaacactctcctaagaactggcagctcctggcagtc 711
 Db 104778 ndnnccgcbggcgccgcsnnndnnccgcbggcgccgcccgcggcsnnndnnccgcbggc 104831

RESULT 4

V58939
 ID V58939 standard; DNA; 9960 BP.

XX AC V58939;

XX DT 15-FEB-1999 (first entry)

XX DE Mycobacterium smegmatis embCAB operon.

XX KW Drug resistance; antibiotic resistance; antimycobacterial;

XX KW ethambutol; embCAB operon; infection; vaccine; therapy; ds.

XX OS Mycobacterium smegmatis.

XX PN WO9841533-A1.

XX PD 24-SEP-1998.

XX PF 16-MAR-1998; 98WO-U805128.

XX PR 20-MAR-1997; 97US-0822586.

XX PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

XX PI Jacobs WR, Musser J, Telenti A;

XX DR WPI; 1998-521160/44.

XX DR P-PSDB; W73055-57.

XX PT Wild type and mutated sequences of Mycobacterium embCAB operon -

XX PT useful to e.g. identify ethambutol-resistant mycobacterial strains

XX PT and produce antisense sequences to treat mycobacterial infections

XX PS Disclosure: Fig 6A-I; 62pp; English.

XX This is the DNA sequence of the Mycobacterium smegmatis embCAB

CC operon, which determines resistance to the antimycobacterial drug

CC ethambutol (EMB). It includes the embA, embC and embB genes that

CC encode proteins (see W73052-54) which are the target of action of

CC M. smegmatis for EMB. To identify genes conferring EMB resistance,

CC a genomic library from a high level EMB-resistant mutant of M.

CC smegmatis was introduced into wild-type M. smegmatis mc2155. Four

CC overlapping cosnids were identified which conferred a resistant

CC phenotype. The minimum size fragment capable of conferring EMB

CC resistance was 9 kb (pIMW99). pIMW99 plus 7 kb upstream M.

CC smegmatis sequence was sequenced revealing 3 homologous open

CC reading frames (embC, embB, embA) and 4 additional potential coding

CC regions. Wild-type and mutated embCAB nucleic acid sequences are

CC useful as probes used in the diagnosis of drug-resistant

CC mycobacteria or to determine the susceptibility of mycobacteria to

CC EMB. The nucleic acids are also useful in the treatment of

CC mycobacterial infections; anti-DNA or anti-RNA sequences can be

CC administered to inhibit embCAB operon mRNA activity (claimed). The

CC invention additionally provides for the use of embCAB operon

CC nucleic acid sequences as vaccines, or to improve existing

CC vaccines.

XX Sequence 9960 BP; 1389 A; 3592 C; 3360 G; 1618 T; 1 other;

Query Match

Best Local Similarity 3.0%; Score 46.4; DB 19; Length 9960;

Matches 170; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 831 gcccacatcgccatgtgagctcagatggcagcagcatcctgcagaccatccctgccacaaacc 890

Db 2385 ggcctccggcgcgcgtgacctcgtcgggttcggcgacggcatccctcggatgtctc 2444

QY 891 cctgtctccagggtgctcctggagggagacagagagggcttctacctctaccagatggaaa 950

Db 2445 cgcggaccctcctcgtggagcagccgcgcagcaactcgcgacagcagcactcgcggcgt 2504

QY 951 gaccac 1010

Db 2505 cgtcaccggcaccgaggtcggcagcaggaaggcggcagcagcagcagcagcagcagcagc 2564

QY 1011 gtcagagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1070

Db 2565 atcccgcgcgccctgcgctacggcgtgacccggccacacacacacacacacacacacacac 2624

QY 1071 ctgtgtcagagac 1130

Db 2625 gcgttcgggac 2684

QY 1131 ctgctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1190

Db 2685 cgaaccagcgccgc 2744

QY 1191 caagggtcggaggcc 1206

Db 2745 cgaggtgcagtgggcc 2760

RESULT 5

X53491/C

ID X53491 standard; DNA; 114955 BP.

XX AC X53491;

XX DT 05-JUL-1999 (first entry)

XX DE Human adenosine A1 receptor antisense oligonucleotide fragment.

XX KW Antisense oligonucleotide; multiple target; antisense treatment;

XX KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 XX
 OS Synthetic.
 XX
 PN W09913886-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 17-SEP-1998; 98WO-US19419.
 XX
 PR 09-JUN-1998; 98US-0093972.
 PR 17-SEP-1997; 97US-0059160.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;
 XX
 DR WPI; 1999-229400/19.
 XX
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 XX
 PS Disclosure; Page 37; 120pp; English.
 XX
 CC The specification describes antisense oligonucleotides (X52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of RNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences X5272-74. These multiple target
 CC oligonucleotides (specifically X55180-271) can be used for the antisense
 CC treatment of diseases and conditions. Typical diseases and conditions
 CC are those associated with impaired respiration and inflammation,
 CC including lung diseases, pulmonary vasoconstriction, inflammation,
 CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
 CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
 CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
 CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
 CC metastases, as well as all types of cancers which may metastasize or have
 CC metastasized to the lungs, including breast and prostate cancer.
 XX
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Db 104875 HNNSVGGCCVCGCGNNHNNSCCVCGCGNNHNNSCCVCGCGNNHNNSC 104816
 QY 240 tggggactttctactatcctgccaatctggagcccaagagcagcagtgccgc 299
 Db 104815 CCVGGCCVCGCGNNHNNSCCVCGCGNNHNNSCCVCGCGNNHNNSC 104756
 QY 300 gctgctgctctcccgaggccgaagagtgccaacagcagctcttcgagcggtccca 359
 Db 104755 NSGCGCCVCGCGNNHNNSCCVCGCGNNHNNSCCVCGCGNNHNNSC 104696
 QY 360 actcagcgcagctgccaagctgccaatctcttcagcc 400
 Db 104695 CCVCGCGNNHNNSCCVCGCGNNHNNSCCVCGCGNNHNNSC 104655
 RESULT 6
 Z56005/c
 ID Z56005 standard; DNA; 1693 BP.
 XX
 AC Z56005;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Contig 004 from cosmid PKOS023-27 from Streptomyces venezuelae.
 XX
 KW Narbonolide polyketide synthase; PKS; cosmid PKOS023-27; contig 004;
 KW ketolide; SAM synthase; S-adenosylmethionine synthase; hydroxylase;
 KW picromycin; antibiotic production; narbomycin; ds.
 XX
 OS Streptomyces venezuelae.
 XX
 FH Key Location/Qualifiers
 FT CDS complement (694..1692)
 FT /tag= a
 FT /product= SAM synthase
 FT /note= "Partial S-adenosylmethionine synthase"
 FT CDS complement (3..692)
 FT /tag= b
 FT /product= ORF16 product
 FT /note= "M. tuberculosis cbhK homologous partial protein"
 XX
 PN W09961599-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 27-MAY-1999; 99WO-US11814.
 XX
 PR 28-MAY-1998; 98US-0087080.
 PR 28-AUG-1998; 98US-0141908.
 PR 22-SEP-1998; 98US-0100880.
 PR 08-FEB-1999; 99US-0119139.
 XX
 PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX
 PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
 XX
 DR WPI; 2000-072613/06.
 DR P-PSDB; Y67216, Y67217.
 XX
 PT New recombinant DNA encoding a domain of narbonolide polyketide
 PT synthase, for production of ketolide antibiotics -
 XX
 PS Disclosure; Page 38; 98pp; English.
 XX
 CC This is contig 004 from the recombinant cosmid PKOS023-27 DNA sequence
 CC (see Z56001) which contains a Streptomyces venezuelae DNA insert. The
 CC cosmid contains open reading frames which encode the various modules of
 CC the narbonolide polyketide synthase (PKS). The invention relates to
 CC recombinant DNA containing a coding sequence for a narbonolide PKS.
 CC Polyketides are compounds synthesised from 2-carbon units through a
 CC series of condensations and subsequent modifications. Modular PKSs are
 CC responsible for the production of many antibiotics including picromycin.
 CC The narbonolide PKS consists of a loading module, six extender modules,

Query Match 2.8%; Score 43; DB 20; Length 114955;
 Best Local Similarity 31.9%; Pred. No. 2.8;
 Matches 128; Conservative 46; Mismatches 225; Indels 2; Gaps 1;
 QY 2 gcgaggctccctgctgctggcggtggcccggtggggggagacagggaagagccgcg 61
 Db 105055 GCGCGCGCGCCVCGCGNNHNNSCCVCGCGNNHNNSCCVCGCGNNHNNSC 104996
 QY 62 cccctggcgccggcagtcaggtcgcagcgcctcgaagagcaatgcgtgcaccccgcc 121
 Db 104995 GCGCGCGCGCGCGCGCCVCGCGNNHNNSCCVCGCGNNHNNSCCVCGCGCG 104936
 QY 122 tctcgtgagtcctccctgcgtgcgggacctgctcccgccagcgcag--ctgctcg 179
 Db 104935 CVCNNHNNSCGCGCGCGCGCGCGCCVCGCGNNHNNSCCVCGCGNNHNNSC 104876
 QY 180 agaggtggccattctcgccggcgccggcgagcgcccgccgggtcccgcggtc 239

XX
DA (MTNE) UNIV MINNESOTA

AC 256004;

DT 23-MAR-2000 (first entry)
XX
DE Contig 003 from cosmid PKOS023-27 from Streptomyces venezuelae.
XX
KW Narbonolide polyketide synthase; PKS; cosmid PKOS023-27; contig 003;
KW ketolide; dNDP glucose synthase; dehydratase; picromycin; narbomycin;
KW antibiotic production; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 104..982
FT CDS /*tag= a
FT /*product= dNDP_glucose_synthase
FT /*note= "glucose-1-phosphate thymidyl transferase"
FT CDS 1114..2127
FT /*tag= b
FT /*product= dNDP_glucose4_6_dehydratase
FT CDS 2124..3263
FT /*tag= c
FT /*product= PICCI
FT /*transl_except= (Pos:2874..2876, aa:Xaa)
FT /*note= "Xaa = Unknown"
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
XX 27-MAY-1999; 99WO-US11814.
XX
XX 28-MAY-1998; 98US-0087080.
XX 28-AUG-1998; 98US-0141908.
XX 22-SEP-1998; 98US-0100880.
XX 08-FEB-1999; 98US-0119139.
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX WPI; 2000-072618/06.
XX P-PSDB; Y67206, Y67214, Y67215.
XX
XX New recombinant DNA encoding a domain of narbonolide polyketide
XX synthase, for production of ketolide antibiotics .
XX
XX Disclosure: Page 37-38; 98pp; English.

XX
XX This is contig 003 from the recombinant cosmid PKOS023-27 DNA sequence
XX (see Z56001) which contains a Streptomyces venezuelae DNA insert. The
XX cosmid contains open reading frames which encode the various modules of
XX the narbonolide polyketide synthase (PKS). The invention relates to
XX recombinant DNA containing a coding sequence for a narbonolide PKS.
XX Polyketides are compounds synthesised from 2-carbon units through a
XX series of condensations and subsequent modifications. Modular PKSs are
XX responsible for the production of many antibiotics including picromycin.
XX The narbonolide PKS consists of a loading module, six extender modules,
XX and two thioester domains. Four proteins make up the narbonolide PKS
XX (PICAI, PICAI, PICAI and PICAI). PICAI includes the loading module
XX and extender modules 1 and 2, PICAI includes extender modules 3 and 4,
XX PICAI includes extender module 5 and PICAI includes extender module 6
XX and a type II thioesterase domain. The second type II thioesterase domain
XX is found on the PICB protein. The nucleotide sequences encoding all of
XX these proteins can be isolated in recombinant form from the recombinant
XX cosmid PKOS023-27. Narbonolide is desaminylated in S. venezuelae to
XX yield narbomycin, the desaminyl transferase enzyme is required for this
XX conversion, and the desaminase biosynthetic genes are also found in
XX cosmid PKOS023-27. The recombinant DNA of the invention is used to
XX express, in transformed cells, narbonolide (or its derivatives) or other
XX ketolides (particularly hybrids), which may then be converted (e.g. by
XX other enzymes recombinantly expressed in the same hosts) to polyketide
XX antibiotics or their intermediates. The antibiotics are useful in human
XX or veterinary medicine.

SQ Sequence 3291 BP; 521 A; 1289 C; 1044 G; 436 T; 1 other;
Query Match 2.7%; Score 41.6; DB 21; Length 3291;
Best Local Similarity 46.3%; Pred. No. 2.7;
Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
Qy 85 ctgcagcgccctagagagcgaatcgctcgagcccccccggtgtccgtgagtcctcccttcgctg 144
Db 3247 CCGCTCGGCCCAATTCGGCGACGGCTCGATCACCGCAGGGCTGCGGGGCTCCAGGTG 3188
Qy 145 cgggacctgtgccccgcacagcgagctcttcgagagtggtgcccattctcggcgggcg 204
Db 3187 CGGCCCATGCGCAGGCTGAGGACCTGCGCGGGGAGAGTGCACGGGTACCGGGTAGTCGT 3128
Qy 205 gcggcgagggcgccccgggggtcccgcggtctctggggactttctactatctactaactg 264
Db 3127 TTCGGCGGGTTCCTCGCCCGGTAGCGGGGAGAGTGCACGGGTACCGGGTAGTCGT 3068
Qy 265 gccaatctgagggccaaagacagcagcagtggtgccccgctgctcccccggggcggaag 324
Db 3067 GAGGTGTGATGCGCGGGGCGGCTGCTGCGCAGCTGCTGCGGGCGCTCGGTGCG 3008
Qy 325 agtgcacacgagcagctcttcggcggtcccgagactcagcgagcagctgggcaaa 380
Db 3007 CACGTTGAAGAGGTGCCAGACCGGGTCTGGGTGCGGGCGGTACCGGCAGGCCGA 2952
RESULT 9
Z87284
ID Z87284 standard; DNA; 12441 BP.
XX
AC Z87284;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
KW Desosamine biosynthesis; macrolide; polyketide; picromycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX WPI; 2000-150679/14.
DR P-PSDB; Y77179.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin -
XX
XX Claim 2; Page 281-287; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the erylC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesise methymycin,

CC engineer PHA monomer synthases or to prepare biologically active agents,
CC as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polypeptides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 32.
XX
SQ Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

Query Match 2.7%; Score 41.6; DB 21; Length 13613;
Best Local Similarity 46.3%; Pred. No. 3.6;
Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 85 ctgacgogctagaagaatgctgaccccccggctgctcgtgagtcctcccttcgtcg 144
Db 6845 ccgctggccattcgccgagcgtcgatcaaccgcagcgtcgccgctccagggtg 6904
QY 145 cgggaactgtgcccgcacacgcagcgtcttcgagaggtgcccatttcggcgcg 204
Db 6905 cgggccgagtcgagcgtgagacgtgcccgcgagcgtctcggcgagcgagcc 6964
QY 205 gcggcgagggcgcccccgggggtcccgcggtctctggggagctttactcatctac 264
Db 6965 ttccggcggtgctcccgcggtagcgggcgagaggtgcaagggtacgggttagtgc 7024
QY 265 gccaatctggagcgaagcagcagcaggtggtcgcgctgctcctcccgggcgcaagg 324
Db 7025 gagggtgctgatcccgggcggtcgaggtggtcgagcgtcgtcggtcggtcg 7084
QY 325 agtgcaacgacagctctccggcggggtccagactcagcgacagctggccaa 380
Db 7085 caggtgaagaggtgccagacgggtcggtgctggcggtcagcgagggccga 7140

RESULT 11
X08910
ID X08910 standard; cDNA; 1582 BP.

AC X08910;
XX
DT 27-APR-1999 (first entry)
DE Human FADD protein coding sequence.

XX FIP; FADD interacting protein; FADD; Fas-associated protein with a
KW novel death domain; cell death; apoptosis; Alzheimer's disease;
KW Acquired Immune Deficiency Syndrome; AIDS; muscular dystrophy;
KW amyotrophic lateral sclerosis; virus; bacteria; fungus; mycoplasma;
KW protozoa; neoplasia; dysplasia; hyperplasia; ds.
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FH 70..696
FT CDS /tag= a
FT FT /label= Human_FADD_protein

XX W09000499-A1.
XX
XX
XX 07-JAN-1999.
XX
XX 26-JUN-1998; 98W0-0513320.
XX
XX 03-JUN-1998; 98US-0087886.
XX
XX 26-JUN-1997; 97US-0050792.
XX
XX (CHIR) CHIRON CORP.

XX Chen TT, Williams LT;
PI
XX WPI; 1999-095745/08.
DR P-FSD; W96134.
DR
XX New FADD (Fas-associated protein with a novel death
PT domain)-Interacting Protein - useful for inducing or preventing
PI apoptosis in a cell, to aid in controlling apoptosis-related
PI diseases
XX
PS Disclosure; Page 45-46; 58pp; English.

XX An epitope of human FADD (Fas-associated protein with a novel
CC death domain)-Interacting protein (FIP protein) comprising amino
CC acids 348-727 of the protein described in W96153, can be used to
CC induce or prevent apoptosis in a cell. Specifically, decreasing the
CC levels of FIP348-727 prevents apoptosis. This is useful in cells
CC which are dying prematurely, eg: Alzheimer's disease, Acquired
CC Immune Deficiency Syndrome (AIDS), muscular dystrophy, amyotrophic
CC lateral sclerosis (and other muscle wasting diseases), autoimmune
CC diseases, and diseases where cells are infected with a pathogen
CC (virus, bacteria, fungus, mycoplasma or protozoa). Increasing the
CC levels of FIP 348-727 induces apoptosis which is useful in cells
CC suffering from neoplasias, dysplasias, hyperplasias, or their
CC symptoms. Purified and isolated FIP subgenomic polynucleotides are
CC useful as primers to obtain more copies of the nucleotides, and as
CC probes that identify wild-type or mutant coding sequences. They are
CC also useful for expressing FIP mRNA, proteins or fusion proteins,
CC and in the generation of FIP antisense oligonucleotides and
CC ribozymes. They are also useful in expression constructs and in
CC gene delivery vehicles (optionally in combination with a condensing
CC agent) that deliver FIP mRNA or oligonucleotides, FIP proteins
CC (including variants), FIP-specific ribozymes or single-chain
CC antibodies into eukaryotic cells. This is the human FADD protein
CC gene. Human FIP protein binds to amino acids 1-110 of the human
CC FADD protein given in W96154.
XX
SQ Sequence 1582 BP; 344 A; 433 C; 483 G; 322 T; 0 other;

Query Match 2.7%; Score 41.4; DB 20; Length 1582;
Best Local Similarity 48.9%; Pred. No. 2.5;
Matches 111; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 33 gtggggcgacagtgggaaggcccgccctggcgggcgagtcaggatgctgcagcg 92
Db 156 gcgcgtggcgacgcaagctggagcgtgcagagcgccctgctctctccatgct 215
QY 93 cctagaagagcaatgctgcaccccggtgctcgtgagtcctcctgcggggacct 152
Db 216 gctggagcagacacctggagccggcgacacacgagctcctgcgagctgctgcctc 275
QY 153 gctgccccgcacagcagctgcttcgagaggtgcccattctcgcgggcgcgcg 212
Db 276 cctggcgccacgacctgctgcggcggtgcgacgacttcgagggggcgggcg 335
QY 213 agggggccccgggtcccgcggtctctggggactttctactatct 259
Db 336 ggcgcgcctgggggaagacctgtgctgagcatttaacgtcatat 382

RESULT 12
T39397
ID T39397 standard; cDNA; 1642 BP.

XX T39397;
XX
XX 22-FEB-1997 (first entry)
XX
XX FADD (Fas-associating protein with novel death domain) gene.
DE
XX Human; FADD; Fas-associating protein with novel death domain;
KW

KW apoptosis; Fas receptor; death domain; gene therapy; antibody;
 KW immunosay; drug screening; diagnostic; AIDS; antiinflammatory;
 KW antitumour; cerebroprotective; neuroprotective; ss.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH 5'UTR 1..129 /*tag= a
 FT misc_feature 4..6 /*tag= b
 FT /*tag= "In-frame stop codon"
 FT CDS 130..756
 FT /*tag= c
 FT /*product= FADD protein
 FT misc_feature 133..501
 FT /*tag= d
 FT /*note= "Encodes N-terminal half, inducing apoptosis
 but not binding Fas receptor"
 FT misc_feature 198 /*tag= e
 FT /*note= "Clone 15 start point"
 FT misc_feature 249 /*tag= f
 FT /*tag= "Clone 8 start point"
 FT misc_feature 450..660
 FT /*tag= g
 FT /*note= "Region encoding death domain"
 FT 3'UTR 757..1642
 FT /*tag= h
 FT polyA_signal 1636..1641
 FT /*tag= i
 XX WO9631603-A2.
 PN 10-OCT-1996.
 XX 28-FEB-1996; 96WO-US02857.
 XX 18-MAY-1995; 95US-0443982.
 PR 03-APR-1995; 95US-0416379.
 XX (UNMI) UNIV MICHIGAN.
 PA Dixit VM, O'Rourke K;
 PI WPI; 1996-465026/46.
 DR P-PSDB; W03653.
 XX FADD protein that binds to cytoplasmic region of Fas receptor - for
 PT identifying inhibitors of Fas-associated apoptosis useful for
 PT treating e.g. AIDS, leukemia, stroke, etc

Example 1; Fig 2A-B; 96pp; English.

XX The sequence encodes FADD (Fas-associating protein with novel death
 CC domain), which binds the cytoplasmic region of a Fas receptor, and
 CC modulates apoptosis induced by activation of the receptor by ligand
 CC binding. The cDNA has been isolated using a yeast two-hybrid system
 CC which screens for proteins interacting with the Fas cytoplasmic
 CC domain. A GAL4 DNA-binding domain has been fused to the human Fas
 CC antigen cytoplasmic tail to form a bait plasmid, which is used with
 CC the GAL4-activation domain, to co-transform yeast cells.
 CC Overlapping clones 8 and 15 have been isolated, and the full-length
 CC cDNA has been isolated from a human umbilical vein endothelial cell
 CC (HUVEC) library using clone 15 as a probe. The cDNA has an
 CC in-frame stop codon 130 bp upstream of the Initiator Met. The
 CC encoded protein contains a death domain, with interacts with the
 CC death domain of Fas. The DNA may be used in gene therapy, and the
 CC protein or a corresponding antibody may be used to screen for
 CC agents modulating FADD pathway cellular functions and Fas-associated
 CC apoptosis, for use in therapy of e.g. AIDS, inflammation,

CC Leukaemia, myocardial infarction, degenerative disease, etc.
 XX
 SQ Sequence 1642 BP; 354 A; 448 C; 508 G; 332 T; 0 other;

Query Match 2.7%; Score 41.4; DB 17; Length 1642;
 Best Local Similarity 48.9%; Pred. No. 2.6;
 Matches 111; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 33 gtgggggagagtagtgggaagagggccgcgcctggggcgagtcaggtgctgcagcg 92
 Db 216 ggcgctgggagcagcaagctggagcgtgcagagcgctacacctctctccatgct 275
 QY 93 cctagaagagcaatgctgcaccccggtgctcgtgagtcctcctcctcgtgcgggacct 152
 Db 276 gctggagcagracgacctggagcccgggcacacgagctcctgcgagctgcgctc 335
 QY 153 gctgcccgcacagcgagctgcttcgagaggtggccattctcggggcgcgcg 212
 Db 336 cctggcgcgacgacctgctgcggcgctgcgagcttcgagcgggcgcgcgcg 395
 QY 213 agcgcgccccggggtcccgcggtctctgggacctttctactatct 259
 Db 396 ggccgcgctggggaagagacctgtgtgcagcatttaacgtcatat 442

RESULT 13

XX81941
 ID X81941 standard; cDNA; 2644 BP.

XX AC X81941;

XX DT 09-SEP-1999 (first entry)

XX DE cDNA encoding a protein involved in eliciting a signal in HH-PTC.

XX KW Human homologue; Drosophila fused gene; intracellular signal;
 KW human hedgehog-patched protein; HH-PTC; RNA transcription;
 KW RNA translation; cancer diagnostic; prophylaxis; karyotyping analysis;
 KW cancer; embryonic repair; tissue repair; wound healing;
 KW neurodegenerative disease; testicular function; ss.

XX OS Homo sapiens.

XX PN WO9932609-A1.

XX PD 01-JUL-1999.

XX PF 18-DEC-1998; 98WO-S02384.

XX PR 26-JUN-1998; 98SE-0002292.

XX PR 19-DEC-1997; 97SE-0004788.

XX PA (KARO-) KAROLINSKA INNOVATIONS AB.

XX PI Toftgard R, Zaphiropoulos PG;

XX DR WPI; 1999-418918/35.

XX DR P-PSDB; Y23755.

XX PT Human homologue of the Drosophila fused gene

XX PS Claim 11; Page 74-77; 81pp; English.

XX The present sequence encodes a human homologue to molecules associated
 CC with the Drosophila fused gene. The protein is involved in eliciting an
 CC intracellular signal in the human hedgehog-patched (HH-PTC). The
 CC protein, its antibodies and pharmaceutical compositions comprising them
 CC can be used as medicaments. The presence of a fused gene, cDNA, mRNA,
 CC protein or subsequences of these in a biological sample are useful,
 CC e.g. as a marker to assess in vivo and/or in situ RNA transcription
 CC and/or translation, in cancer diagnostics, in prophylaxis, etc. The
 CC polynucleotide sequence can also be used to derive probes for the

PI Elshourbagy N, Lane PA, Tsui P, Vawter L;

XX Tandem repeat polymorphism; B2 bradykinin receptor; diagnosis
KW myocardial infarction; coronary heart disease; cardiomyopathy

$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = \frac{\partial L}{\partial x}$

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 12:41:38 ; Search time 1490.04 Seconds
(without alignments)
5900.480 Million cell updates/sec

Title: us-09-434-708-3
Perfect score: 1422
Sequence: 1 atggtctgtgggtggccccc.....ccagagaccctgcccgcc 1422

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- FST:*
- 1: gb_est1.*
 - 2: gb_est2.*
 - 3: gb_est3.*
 - 4: gb_est4.*
 - 5: gb_est5.*
 - 6: gb_est6.*
 - 7: gb_est7.*
 - 8: gb_est8.*
 - 9: gb_est9.*
 - 10: gb_est10.*
 - 11: gb_est11.*
 - 12: gb_est12.*
 - 13: gb_est13.*
 - 14: gb_est14.*
 - 15: gb_est15.*
 - 16: gb_est16.*
 - 17: gb_est17.*
 - 18: gb_est18.*
 - 19: gb_est19.*
 - 20: gb_est20.*
 - 21: gb_est21.*
 - 22: gb_est22.*
 - 23: gb_est23.*
 - 24: gb_est24.*
 - 25: gb_est25.*
 - 26: gb_est26.*
 - 27: gb_est27.*
 - 28: gb_est28.*
 - 29: gb_est29.*
 - 30: gb_est30.*
 - 31: gb_est31.*
 - 32: gb_est32.*
 - 33: gb_est33.*
 - 34: gb_est34.*
 - 35: gb_est35.*
 - 36: gb_est36.*
 - 37: gb_est37.*
 - 38: gb_est38.*
 - 39: gb_est39.*
 - 40: gb_est40.*
 - 41: em_estba.*
 - 42: em_estfun.*
 - 43: em_esthum1.*

- 44: em_esthum2.*
- 45: em_esthum3.*
- 46: em_esthum4.*
- 47: em_esthum5.*
- 48: em_esthum6.*
- 49: em_esthum7.*
- 50: em_esthum8.*
- 51: em_esthum9.*
- 52: em_esthum10.*
- 53: em_esthum11.*
- 54: em_esthum12.*
- 55: em_esthum13.*
- 56: em_esthum14.*
- 57: em_esthum15.*
- 58: em_esthum16.*
- 59: em_esthum17.*
- 60: em_esthum18.*
- 61: em_esthum19.*
- 62: em_esthum20.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estom.*
- 68: em_estov1.*
- 69: em_estov2.*
- 70: em_estpl1.*
- 71: em_estpl2.*
- 72: em_estpl3.*
- 73: em_estpl4.*
- 74: em_estpl5.*
- 75: em_estro1.*
- 76: em_estro2.*
- 77: em_estro3.*
- 78: em_estro4.*
- 79: em_estro5.*
- 80: em_estro6.*
- 81: em_estro7.*
- 82: em_estro8.*
- 83: em_estro9.*
- 84: em_estro10.*
- 85: em_estro11.*
- 86: em_estro12.*
- 87: em_estro13.*
- 88: gb_gss1.*
- 89: gb_gss2.*
- 90: gb_gss3.*
- 91: gb_gss4.*
- 92: em_gss1.*
- 93: em_gss2.*
- 94: em_gss3.*
- 95: em_gss4.*
- 96: gb_gss5.*
- 97: gb_gss6.*
- 98: gb_gss7.*
- 99: gb_gss8.*
- 100: gb_gss9.*
- 101: em_gss5.*
- 102: em_gss6.*
- 103: em_gss7.*
- 104: em_gss8.*
- 105: em_gss9.*
- 106: em_gss10.*
- 107: em_gss11.*
- 108: gb_gss10.*
- 109: gb_gss11.*
- 110: em_gss12.*
- 111: gb_gss12.*
- 112: gb_gss13.*
- 113: gb_gss14.*
- 114: gb_gss15.*
- 115: gb_gss16.*
- 116: gb_gss17.*

Db 513 GCAGCTCTACTGGGCCATGGACTCCACATNTGAGCTCTGCAAGATCTGTGCTGAGAGCAA 454
QY 1074 caaggatgtgaagattgagccgtggggcaccctgtctgcagctgtctgctgctgctgctg 1133
Db 453 CRAGGATGTGAGATGTGACCTGTGGGACACCTGTCTGACCTGTCTGCTGCTGCTGCTG 394
QY 1134 ggaagcactggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1193
Db 393 GCAGCACTGGGACAGCAGACCTGCCCCCTCTGCGGCTCGGAGATCAAGGGCTGGAGGC 334
QY 1194 cgtgaatattaccagttccaggtcagcagcagcagcagcagcagcagcagcagcagcagc 1253
Db 333 CGTGAATCTACCACTTCACGGTCAGGCTACTGTGAGGACTCAGGGAACAGCAGTGA 274
QY 1254 ccagggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1313
Db 273 CCAGGAAGCAGGAGTGTGAGCTGGGCGAGGTGCCCCCTTCGGCTCCCTCCATTGCCGCC 214
QY 1314 agggccagatctgccccccaggaagccagaaatgccccagccaggaagtgagactcctaaa 1373
Db 213 AGGCCAGATCTGCCCCAGGAAGCCAGAAATGCCAGCCGAAAGTAGAGACTCCTAAA 154
QY 1374 ggggaactccctccagctgctgggacccagcagcagcagcagcagcagcagcagcagc 1422
Db 153 GGGGAACCTCCCTCCAGCTGCGCTGGGACCCAGGACCTGCCCGGCC 105

RESULT 2
BE407909
LOCUS BE407909 695 bp mRNA EST 21-JUL-2000
DEFINITION 601299249F1 N1H_MGC_21 Homo sapiens cDNA clone IMAGE:3629101 5',
mRNA sequence.
ACCESSION BE407909
VERSION BE407909.1 GI:9344359
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 695)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM315 row: b column: 14
High quality sequence stop: 686.
Location/Qualifiers
1. .695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629101"
/clone_lib="N1H_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

118 a 241 c 206 g 130 t
BASE COUNT
ORIGIN

FEATURES
source

Query Match 39.0%; Score 554; DB 35; Length 695;
Best Local Similarity 94.0%; Pred. No. 4.8e-118;
Matches 643; Conservative 0; Mismatches 30; Indels 11; Gaps 6;
QY 63 agtcagatgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 122
Db 1 AGTCAGATGCTGAGCGCCTAGAGAGCAATGCTGCAGCCCGGCTGTCCTGAGTCC 50
QY 123 cccttcctcgggacacgtgctcccccacagcagcagcagcagcagcagcagcagcagcagcagc 181
Db 61 CCCCTGCTGGGAGACCTGTGTCGCCCCACAGCGCAGCTGCTTCGACGAGGTGCCCAT 120
QY 182 ctcggcggggggcggcggg--aggcggcccccgggggtcccccgggggtctctgggagctttct 239
Db 121 CTCGGCGGGCGCGCGGGGAGGCGCGGGTGGTCCCGGGGCTCTGGGAGCTTCT 180
QY 240 actcat--ctaccctggccaatctgagggcccaagcagcagcagcagcagcagcagcagcagc 297
Db 181 ACTCATCTATCTCTGCGCAATCTGGAGGCCAAGAGGAGGAGGTGGCGGCGCTGCTGCT 240
QY 298 cccggggggcgaagagtgccaac----gacgagctcttcgggggggggtcccccagcagcagcagc 353
Db 241 CCCCAGGCGGAGGATGTGCACACGACTGATGCTCTTCGGGCGGGCTCCAGACTCAG 300
QY 354 ggcacagctggccaa--gctggccatcttccagccacatgcagcagcagcagcagcagcagcagc 412
Db 301 CGGACTGCTGCCCAATGCTGCCCATCTCTCAGCCACATGCACGCGAGAGCTGCACGCAC 360
QY 413 tctcccccggggcgaagctgtgacacatgtaccagctccacaaaggcccccggccacaca 472
Db 361 TCTTCCCGGGGGAAGTACTGTGACACATGTACACAGCTCACCAGGCGCGCGCCACACA 420
QY 473 cctctggagggaaagtggggagccgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 532
Db 421 CCTTCTGAGGGAAGTTGGGAGCCCGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 533 tctcggcacctgcacccctgtgaacacagctgcacagcagcagcagcagcagcagcagcagcagc 592
Db 481 TCTTGGGACCTGCGCCACCTCTGTGGAACACAGCTGCACGCGCTGCGCTGCGCTGCGCTG 540
QY 593 ttgacctcacctgcagcgggcagcagctgtccatcttcctgagcttcgagcttccaccaggtct 652
Db 541 TTGACCTCACCTGCAGCGGGCAGCTGTCCATCTTCGAGTTCGACGCTTCCACAGGCTCT 600
QY 653 ttcagc 712
Db 601 TTGAGCATGTCGCAACACTCTCTCAAGAACTGGCAGCTCTTGGCAGTCAACCA-CCAGGCT 659
QY 713 acatggccttcctccacctatgatg 736
Db 660 ACATGGCTTCACATATGATGAGG 683

RESULT 3
BE464528
LOCUS BE464528 522 bp mRNA EST 27-JUL-2000
DEFINITION hv84h11.k1 NCI-CCAP_Lu24 Homo sapiens cDNA clone IMAGE:3180165 3',
similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2. ; mRNA
sequence.
ACCESSION BE464528
VERSION BE464528.1 GI:9510303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 522)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 508.

FEATURES

source
 1. .522
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3180165"
 /clone_lib="NCI-CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 117 a 168 c 153 g 84 t
 ORIGIN

Query Match 36.3%; Score 515.8; DB 35; Length 522;
 Best Local Similarity 99.6%; Pred. No. 3.1e-109;
 Matches 517; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 866 ctgccacaacccctgtccagggtctctggagggaagaaggcgtcttaacctt 925
 DB 4 CCGCCAAACACCCCTGCTCCAGGTGCTCTGGAGGACAGAGGAGCGTCTTACCTT 63
 QY 926 accagatggaagaccacacacccagacctgactgagctcgccaggcagacccaccg 985
 DB 64 ACCAGATGAAGACCCACACCCAGACCTGACTGAGCTCGCCAGGAGACCCACG 123
 QY 986 agcgatccactgtcagaggagcagctgcagctctacttggccatggactccacattg 1045
 DB 124 AGCGCATCCAGTGCAGAGGAGCAGCTGCAGCTCTACTGGGCATGGACTCCACATTG 183
 QY 1046 agctctcaagatctgtctgagagcaacaagatgtgaagattgagcgtcgccgacc 1105
 DB 184 AGCTCTCAGATCTGTCTGAGAGCAACACAGGATGTGAGATTGAGCGTGGCGCAC 243
 QY 1106 tgcctcagactgctgctggctggcagcaactcgacacgacccagacctgccctctt 1165
 DB 244 TGCTCTCAGCTGTGCTGGCTGGCTGGCAGCACTCGGACAGCCAGACTGCCCTTCT 303
 QY 1166 ccgctcgagatcaaggctggaggccgtgagtatctaccagttccaggtcaggcta 1225
 DB 304 GCCGCTCGAGATCAAGGGCTGGAGGCGGTGAGTATCTACCATGTTCCACGGTCAGGCTA 363
 QY 1226 ctgctgaggactcagggaacagcagtgaccaggaagcaggagtgaggctggggcagg 1285
 DB 364 CTGCTGAGGACTCAGGGAACAGCAGTGACACAGGAGGAGGTGGAGCTGGGGCAGG 423
 QY 1286 tgccttttggctctcatttgcctccacggccagatctgccccccaggaagcccgaa 1345
 DB 424 TGCCCCCTTCGGCTCTCCATTCGCCCCACGGCCAGATCTGCCCCCATGAAGCCAGAA 483
 QY 1345 atgccacggcgaagtggagactcctaaagggaactccc 1384
 DB 484 ATGCCACGGGAAGTGAGACTCTTAAGGGGAGACTCCC 522

RESULT 4

BE198430 LOCUS 533 bp mRNA EST 26-JUN-2000
 DEFINITION ug78e05.y1 Soares.mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:1548512 5' similar to gb:X57110 PROTO-ONCOGENE C-CBL (HUMAN)>; mRNA sequence.
 ACCESSION BE198430
 VERSION BE198430.1 GI:8710599
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 533)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:953860
 Seq primer: -40RP from Gibco
 High quality sequence stop: 466.
 FEATURES
 Location/Qualifiers
 1. .533
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1548512"
 /clone_lib="Soares.mammary_gland_NMLMG"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

125 a 169 c 133 g 106 t

Query Match

28.9%; Score 411.4; DB 34; Length 533;
 Best Local Similarity 85.7%; Pred. No. 4.3e-85;
 Matches 457; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 634 gacgttccacaggctcttttgaagccatggccacacacctctcaagaactggcagctctg 693
 DB 1 GACGTCTTCACAGGCTCTTTCAGCCGTGGCCACTTTACTGAGGAATTGGCAACTCTG 60
 QY 694 gacgtcaacccccagactacatggcctctcactcatatgatgagttcccaagagcgtctg 753
 DB 61 GCTGTCAACCATCTTGCTGTACATGGCCCTTCTTACCTACCATGAGTTCACCAACAGCCTG 120
 QY 754 caggcctgcaggggacaagccaggcagttacatcttcggcccgctgtactgcctctgggg 813
 DB 121 CAGGCTACAGGGACAAACAGGAGCATATCTTCGGCCAAAGCTGATCCCGCTGGGG 180
 QY 814 cagtgggcccatcggtatgagctcagatggcagcactcctcagaccatccctgccaac 873
 DB 181 CAGTGGGCCCATTTGATACGTGAGTCCGATGGAGAGTCTCTGCAACCATCTCTCTCAAC 240
 QY 874 aaacccctgtcccagggtcctctggaggacagaagcggctcttacctctaccagat 933
 DB 241 AAACCTCTGCTGCAAGTGTCTCTGAAGGGACAAAAGGAGCGCATCTCTCTCTCCCTGAT 300
 QY 934 ggaagaccctcaacccagacctgactgagctggccaggcagaaccccgagcgcatc 993

AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
/W.W. and Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemall.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACGATGACCAT
BACKWARD: GTTTCCTCCATCAGCAG
Plate: 27 row: 1 column: 8
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1. 428
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 480V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos"

BASE COUNT 102 a 133 c 121 g 72 t
ORIGIN

FEATURES source
Query Match 23.6%; Score 336.2; DB 22; Length 428;
Best Local Similarity 89.4%; Pred. No. 1e-67; Indels 0; Gaps 0;
Matches 362; Conservative 0; Mismatches 43;
QY 697 gtaacacacacgggtacatggccttcctccatcatgatgaggtccaaagaggtctgcag 756
Db 24 GGCACACACCGGGTTACATGGCTTCCTCACATAGATGAGGTCCAAACCGCTGTGCAG 83
QY 757 gctgcagaggaacgacgagcaggtacattcttcgcccagctgtactgcctggggcag 816
Db 84 ACTTTCAGAGACAGCAGCAGCAGTACATCTTCGGCCCGCAGTGCATCGCTGGGGAG 143
QY 817 tggggccatggctatgtgagctagatggcagcatcctgcagaccatccctgccacaaa 876
Db 144 TGGGCCATCGGATACGTGAGCTCAGACGGCAGCATCTTCAGACCATCCCTCTCAACAA 203
QY 877 cccctgtcccaggtgctctggaggacagagagcgttctacctaccacagatgga 936
Db 204 CCCCTGTCCAGGTGCTCTGGAGGACAGAGAGGAGGTTTCTACTCTACCCAGATGG 263
QY 937 aagaccacacacccagacctgactgactcggccagggcagaaccccgagcgcacccac 996
Db 264 AAGAACACAAACCGGACCTGACCGAATCTGCCACATGAGACCGCATCAGCATCCAC 323
QY 997 gtctcagagagagctgagctctactggccatctggccatgagctccacatttgagctcgaag 1056
Db 324 GTCTCGAGAGAGAGTGGAGCTTACTGGGCCATGGACTTACGTTTGAGGTCTGCAAG 383
QY 1057 atctgtgctgagagcaacagaggtgtgaagattgagccgtggggg 1101
Db 384 ATCTGTGCCAGAGACACAGGACGTGAAGATAGACCCCTCGGGG 428

RESULT 9
AA113289/c 462 bp mRNA EST 23-DEC-1997
LOCUS zm28c07.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
DEFINITION IMAGE:526956 3' similar to TR:G862411 G862411 CEL-B TRUNCATED FORM
2.; mRNA sequence.

ACCESSION AA113289.
VERSION AA113289.1 GI:1664994
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 462)
REFERENCE Hillier,L., Lennon,G., Becker,M., Donald,M.F., Chiappelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
/M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
/B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E.,
Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Warr,M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
TITLE Contact: Wilson RK
JOURNAL Washington University School of Medicine
MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 367.

FEATURES source
1. 462
/organism="Homo sapiens"
/db_xref="GDB:3918385"
/db_xref="taxon:9606"
/clone="IMAGE:526956"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTGAGTTTCTTTTCTTTT 3"

BASE COUNT 76 a 141 c 137 g 99 t
ORIGIN
Query Match 22.6%; Score 321.6; DB 1; Length 462;
Best Local Similarity 94.6%; Pred. No. 2.5e-64;
Matches 349; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 1054 aagatctgtgatgagcaacaagatgtgaagattgagcgtgcgggcaacctgtctgc 1113
Db 462 AAGATCTCTGTGAGAGCAACAAGATGTGAAGATTGAGCGTCGGGCGACCTTTCTGC 403
QY 1114 agctgctccctggctcctggcagacctcgacag-cacagacctgcccttctccgctg 1172
Db 402 AGCTGCTGCCCTGGGNTGCTGGCAGACTTCGACAGCCAGACCTTCCCTTCGCCGCTG 343
QY 1173 cgagatcaagggctggggagggcgtgagtattaccaggttccacggtcaggctactgtga 1232
Db 342 CGAGATCAAGSGCTGGGAGGCGGTGAGTATCTACAGTTCTACGGTCAAGGTACTGTGA 283
QY 1233 ggaactcagggaaacagcagtgaccaggaaggcaggaggttgaggtggggcagggtgccct 1292
Db 282 GGACTCAGGGAACACAGCAGTACAGGAGGAGGAGTGGAGCTGGGCGAGGTGCCCT 223
QY 1293 ttggctcctcattgccccacacagcagatctgtccccccaggaagccacagaaatgccca 1352

```

Db 222 TTNGGCTCTCCATTGCCCCCAGCGCCAGATCTGCCCCCCAGGAAGCCAGAAATGCCCA 163
QY 1353 gccaagaatgaactcctaaggggaactccctccagctgcctggagccagacc 1412
Db 162 GCCAAGTGAAGTCTCTAAAGGGAAT-CCNCNTCTGCGTGGAGCCCGGAGGACNC 104
QY 1413 tgccccccgc 1421
Db 103 NCTNCCGC 95

RESULT 10
AI368245/c
LOCUS
DEFINITION QW16409.x1 NCI-CGAP_Ut3 Homo sapiens cDNA clone IMAGE:1991273 3'
similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2.; contains
PTR5.t3 MER22 repetitive element.; mRNA sequence.
ACCESSION AI368245
VERSION AI368245.1 GI:4137990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS Contact: Robert Strausberg, Ph.D.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Sequencing by: Greg Lennon, Ph.D.
Cloned by: Washington University Genome Sequencing Center
Cloned distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/obrp/image/image.html
Insert Length: 648 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 359.
FEATURES
source
1..398
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1991273"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="PH103"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"
BASE COUNT 61 a 125 c 129 g 83 t
ORIGIN

Query Match 21.2%; Score 301.4; DB 10; Length 398;
Best Local Similarity 99.7%; Pred. No. 1.1e-59;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1120 tgcttgctgctggcagcactcggagccagccagacctgccctcttcgctggcagatc 1179
Db 398 TGCTTGCTTGCTGGCAGCAGCTGGAGCCAGCAGCTGCCCTTGCGCGCTGCCAGATC 339
QY 1180 aagggtggagggccttgagtattaccagttccaggttcaggctactgctgagactca 1239
Db 338 AAGGGCTGGAGGCGCTGAGTACTTACCAGTTCGCCGGTCAGGCTACTGCTGAGGACTCA 279
QY 1240 gggaacagcagtgaccaggaaggcagggtgaggctggcgaggtgcccccttcgct 1299

```

```

Db 278 GGAACACAGTGTACCCAGAGGAGGAGTGGAGCTGGGAGGTCGCCCTTCGGCT 219
QY 1300 cttccatgccccccagcggcagatctgccccccaggaagcccaagaatgccagccgaaa 1359
Db 218 CTTCCATTGCCCCACGCGCCAGATCTGCCCCAGGAGCCAGAAATGCCAGCCGAAA 159
QY 1360 gtgagactcctaagggaactccctccagctgcctgggagccccagaccctgcccg 1419
Db 158 GTGAGACTCCCTAAGGGGAACTCCCTCCAGCTGGCTGGAGCCCGAGGACCTTGCCCG 99
QY 1420 gcc 1422
Db 98 GGC 96

RESULT 11
F22931
LOCUS
DEFINITION SSC17A01 Porcine small intestine cDNA library Sus scrofa cDNA clone
cl7a01 5', mRNA sequence.
ACCESSION F22931
VERSION F22931.1 GI:2578544
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 491)
AUTHORS Wintero,A.K., Fredholm,M. and Davies,W.
TITLE Evaluation and characterization of a porcine small intestine cDNA
library: analysis of 839 clones
JOURNAL Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE 96327607
COMMENT Contact: A.K. Wintero
Department of Animal Science and Animal Health, Division of Animal
Genetics, The Royal Veterinary and Agricultural University
Bulowsvej 13, 1870 Frederiksberg C, Denmark.
FEATURES
Location/Qualifiers
source
1..491
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="cl7a01"
/clone_lib="Porcine small intestine cDNA library"
/note="directionally cloned cDNA in XLI-blue MRF"
BASE COUNT 116 a 142 c 145 g 79 t 9 others
ORIGIN

Query Match 16.5%; Score 234.6; DB 37; Length 491;
Best Local Similarity 82.0%; Pred. No. 3.3e-44;
Matches 282; Conservative 0; Mismatches 60; Indels 2; Gaps 1;

QY 832 gtgagctcagatggcagcactcctgcagaccatccctgcccaaaacccctgtcccggtg 891
Db 1 GTGAGCTCAGATGGCAGCATCCTGCAGACCATCCTCTCAACAAACCTCTGTTCAGCA 60
QY 892 ctctggaggagacagaaggcgtcttaccttaccttaccagatggaagaccacccca 951
Db 61 CTCCTGGAAGGACAAAAGGAGGCTTCTACTCTTACCTTACCTGATGGGAAGAACCAACCCG 120
QY 952 gacctgactgagctcggccaggcagaaccccccagcgcgtccactgttcagaggagcag 1011
Db 121 GACCTGACGGAGCTCTCGTAAACGGGACCCATCAACGCATCCACGTGTGGAGGAGCAG 180
QY 1012 ctgcagctctactgggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1071
Db 181 CTGAGCTGTACTGGGCGCATGAATCCACATACGAGCTGTGCAAGATCTGTGCCGAGAGA 240
QY 1072 aacagatgtgaagattgagcgtgcccgcagcagcagcagcagcagcagcagcagcagc 1129
Db 241 AACAGGACGTGAAGATTGAGCCATGCGGGGCACTGCTCTGCAAGCCCGCTGCTCGGCTA 300

```

DB 428 TGGACTCCACAATTGAGCTCTGCARAAATCTGTCTGATAGCAACAAGGATGTGAAGATTG 487

QY 1091 agcctgcgggacactgtctgcagctgtgcctgcgtgcctgcct 1132
|||||
DB 488 AGCGTGCGGGACCTGCTCTAGAGTGTTCGTAGGTGCCT 529
|||||

RESULT 13
BE283475 545 bp mRNA EST 13-JUL-2000
LOCUS 601102508F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3495025 5'
DEFINITION mRNA sequence.
ACCESSION BE283475
KEYWORDS BE283475.1 GI:9159750
SOURCE EST.
ORGANISM house mouse.
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC).
TITLE Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8544 row: p column: 02
High quality sequence stop: 542.

FEATURES
Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone="IMAGE:3495025"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary."
/lab_host="DB10B"
/lab_cell="HEp-2"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 97 a 188 c 167 g 93 t
ORIGIN

Query Match 13.2%; Score 187.8; DB 34; Length 545;
Best Local Similarity 70.2%; Pred. No. 2.2e-33;
Matches 384; Conservative 0; Mismatches 142; Indels 21; Gaps 9;

QY 62 cagtcaggatgctcagcgccatagaagagaatcgctgaccccggtgtcctgcgtgagtc 121
|||||
DB 1 CGGTGAAGTTGCTCCAGCGCCTAGAGAGCAATGACGGGATCCCAGGATGTCACGGGCG 60
|||||

QY 122 ccccttcgctcgaggactgctgcccgcacagcgagctcttcgagagtggtgcccatt 181
|||||

DB 61 CCCGTCCTCCGGGAGACTGCTGCCCCGCCACCGGCGAGCTACTTGGAGAGTGGCAAGG 120
|||||

QY 182 ctgcggcgggcgcgcgaggcgcccgcggggtcccgcggtctctgggacattctac 241
|||||

DB 121 CCCGGCGGAGGCGCAGGGAA---GACCCGAGGGTCCCGGTGGCGCCGATGACTTCTGG 177
|||||

QY 242 tcactacctggccaatctggaggccaaagacagcaggtggc-cgcgctctgcctccc 300
|||||

DB 178 CCATCTACCTGGCAATCTGGAGTCAAGGAGGCGAGGTGGCGGAGCTCTCTGCCACCC 237
|||||

/db_xref="taxon:9606"
/clone="IMAGE:2278445"
/clone_lib="NCI_CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

BASE COUNT 37 a 80 c 85 g 56 t
ORIGIN

Query Match 11.5%; Score 164; DB 12; Length 258;
Best Local Similarity 100.0%; Pred. No. 6.4e-28;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1259 aaggcaggagttgagctggggcagggtgccctttcggtctctccattgccccacggc 1318
DB 258 AAGGCAGGAGTTGAGCTGGGGCAGGTGCCCTTTTCGCTCTCTCCATTGCCCCACGGC 199
QY 1319 cagatctgccccccagggaagccagaaatgccagccgaaagtgagactcctaaagggga 1378
DB 198 CAGATCTGCCCCCCCAGGAAGCCAGAAATGCCCCGCGAAAGTGAGACTCTCTAAAGGGGA 139
QY 1379 actccctccagctgctgggagcccccaggaacctggccccggcc 1422
DB 138 ACTCCCTCCAGCTGGGCTGGAGCCCGCAGGACCTGCCCGGCC 95

Search completed: December 7, 2000, 12:41:46
Job time: 24589 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 05:51:57 ; Search time 1490.04 Seconds
(without alignments)
6419.158 Million cell updates/sec

Title: US-09-434-708-1
Perfect score: 1547
Sequence: 1 cgcgaggtcccttgctct.....caaaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

- 44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

117: gb_gss18:.*
118: gb_gss19:.*
119: em_gss13:.*
120: gb_gss20:.*
121: gb_gss21:.*
122: gb_gss22:.*
123: gb_gss23:.*
124: gb_gss24:.*
125: em_gss14:.*
126: em_gss15:.*
127: em_gss16:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	668.8	43.2	751	12	AT660567 we88c03.x
C 2	554	35.8	695	35	BE407909 601299249
C 3	515.8	33.3	522	35	BE464528 hv84h11.x
C 4	512	33.1	530	12	AT690580 tq02a11.x
C 5	411.4	26.6	533	34	BE198430 ug78e05.y
C 6	396.4	25.6	398	10	AT368245 qw16f09.x
C 7	393.6	25.4	462	1	AA113289 zm28c07.s
C 8	370.8	24.0	440	22	AA463722 BP230013B
C 9	356.2	23.0	435	1	AA112513 zm28c07.r
C 10	356.2	21.7	428	22	AA426122 59553 MAR
C 11	258	16.7	258	12	AT758472 t08b03.x
C 12	236	15.3	248	13	AT801220 to86c10.x
C 13	234.6	15.2	491	37	F2931 SSC17A01 PO
C 14	193.2	12.5	529	25	AW857217 RC3-C7030
C 15	187.8	12.1	545	34	BE283475 601102508
C 16	184	11.9	420	25	AW935581 RC3-D7001
C 17	166.6	10.8	251	10	AT344947 t01c03.x
C 18	165	10.7	197	21	AW268319 x95f10.x
C 19	163	10.7	218	10	AT344839 ta99g09.x
C 20	163	10.7	219	21	AW301388 xs76c04.x
C 21	163.4	10.6	207	21	AW301461 xs77c05.x
C 22	162.4	10.5	193	24	AW796557 CM4-UM003
C 23	158	10.2	635	10	AT388785 GH9792.5
C 24	157	10.1	218	21	AW302967 xr87c06.x
C 25	155.4	10.0	218	10	AT345593 tb78d02.x
C 26	152.2	9.8	201	10	AT345656 tb78f04.x
C 27	150.6	9.7	202	21	AW301448 xs77a09.x
C 28	149	9.6	239	21	AW268166 xr93a08.x
C 29	130.4	8.4	469	33	BE061124 OV0-B7C04
C 30	118.2	7.6	379	36	C69305 C69305 fuji
C 31	113.8	7.4	639	14	AL110451 DKF2p434P
C 32	112	7.2	820	18	AV398238 AV398238
C 33	105.4	6.8	548	34	BE201648 fx96d07.x
C 34	105	6.8	360	36	D68728 CELK05AYF
C 35	80.6	5.2	911	123	CNS03VPW
C 36	80.4	5.2	289	40	Z30097
C 37	77	5.0	360	36	C12082 C12082 fuji
C 38	74.2	4.8	179	124	FR0018487
C 39	66.4	4.3	925	121	CNS0091P
C 40	65	4.2	174	124	FR0018465
C 41	64.6	4.2	932	121	CNS0072Q
C 42	59.6	3.9	935	121	CNS006XK
C 43	59.2	3.8	925	121	CNS0091P
C 44	59	3.8	935	121	CNS006XK
C 45	54.2	3.5	932	121	CNS0072Q

ALIGNMENTS

RESULT 1
AT660567/c

LOCUS AT660567 751 bp mRNA EST 18-DEC-1999
DEFINITION we88c03.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone IMAGE:2346244 3' similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2. /, mRNA sequence.
ACCESSION AT660567
VERSION 1 GI:4764137
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 864 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
Location/Qualifiers
1. 751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2346244"
/cclone_lib="Soares_Dieckgraefe_colon_NHCD"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAATCTGAATGGAGCGCGCGCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieckim.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 124 a 222 c 228 g 172 t 5 others
ORIGIN
Query Match 43.2%; Score 668.8; DB 12; Length 751;
Best Local Similarity 95.4%; Pred. No. 3.5e-146;
Matches 718; Conservative 0; Mismatches 32; Indels 3; Gaps 3;
Qy 787 ggcagttacatctccggccagctgtactcgcctggcgagggccatcgactgtg 846
|||||
Db 751 GGCAGTTACATCTTCGCGCCCAACTGACTGCTTTGGGGCAGTGGGCCATCGGCTATGTG 692
Qy 847 agctcagatggcagcatctctgcagaccatccctgcgcacaaacccctgtccc-agtgtct 905
|||||
Db 691 AGCTCAGATGGCAGCATCTTGCAGACATCCTTGGCCCAACAACNCCTGTTCCCAAGTGCT 632
Qy 906 cctggaggagcagagcagcgttcttaccctaccagatggaaagccacacaccaga 965
|||||
Db 631 CTGAAGGGGACAGAGGAGCGGCTTCTA-CTCTACCCAGATGGAAAGACCAACACCCAGNN-573
Qy 966 cctgactgagctcggccaggcagacaaccccgagcgcgcatccacgtgtcagagagcagct 1025
|||||
Db 572 CTTGACTGAGCTCGCCAGCAGAGAA-CCCAGCAGCGCATCCACGTGTCAGAGGAGCAGCT 514
Qy 1026 gcagctctactgggcccagtgactccacatttgagctctgcaagatctgtgctgagagcaca 1085
|||||

513 GCAGCTCTACTGGCCATCGATCCACATNTGAGCTCTGCAAGATCTGTGCTGAGAGCAA 454
1086 caaggtatgtgaagattgagccctgggggacacgtgtctgacgtctgctgctgctgctg 1145
453 CAAGGATGTGAAGATTGAGCCGTGGGACACCTGTCTGTGCTGCTGCTGCTGCTGCTG 394
1146 gcaagactcagacagcagacacacacacacacacacacacacacacacacacacac 1205
393 GCAGCACTGGACAGACACACACCTGCCCTCTGCGCTCGAGATCAAGGGCTGGAGGC 334
1206 cgtgagtattcattcagctcagctcagctcagctcagctcagctcagctcagctcag 1265
333 CGTGAGTATCTACCACTCCAGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAG 274
1266 caggaagcaggagattgagctgagcagcagcagcagcagcagcagcagcagcagc 1325
273 CCAGGAAGCAGGAGATTGAGCTGGGCGAGGTCCTTTCGGCTCCCTCCATTGCCCTC 214
1326 acggccagatctcctcccccaggaagccagaaatgcccagcgcgaagagtagactct 1385
213 ACGCCAGATCTGCCCTCCAGGAAGCCAGAAATGCCAGCGAAGTGAAGTGAAGTGA 154
1386 ggggaactccctccagctgctgctgctgctgctgctgctgctgctgctgctgctg 1445
153 GGGGAATCTCCCTCCAGCTGCGCTGGGACCCAGGACCTGCCCTGCCCTGCCCTGCC 94
1446 gcaccagatgtgctgctcaaggagagcccaagggctggaagggttctgaaacagaa 1505
93 GCACCAGATGTCTCTCAAGGAGACCCCAAGGGCTGGAAGGGGTCTGTAACCGAA 34
1506 taaactgcaagcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1538
33 TAAACTGCCAAGCCTGTCTGTCAAAAAA 1

RESULT 2
BE407909 695 bp mRNA EST 21-JUL-2000
LOCUS 501299249F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629101 5',
DEFINITION mRNA sequence.
ACCESSION BE407909
VERSION BE407909.1 GI:9344359
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 695)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: L1CM315 row: b column: 14
High quality sequence stop: 686.
Location/Qualifiers
1..695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629101"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; vector: pORF7; Site1: XhoI;
Site2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

FEATURES
source
1..695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629101"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; vector: pORF7; Site1: XhoI;
Site2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA Synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 118 a 241 c 206 g 130 t
ORIGIN

Query Match 35.8%; Score 554; DB 35; Length 695;
Best Local Similarity 94.0%; Pred. No. 2.4e-119;
Matches 643; Conservative 0; Mismatches 30; Indels 11; Gaps 6;
Qy 75 agtcagatgctgagcgcctagagcgaatgctgcagccccccggctgctccgtgagtc 134
Db 1 AGTCAGGATGCTGAGCGCCTAGAGAGCAATGCTGACCCCGGGCTGTCCTGAGTCC 60
Qy 135 ccttcgtgctggagacgtgctgcccgcagcagcagcagcagcagcagcagcagc 193
Db 61 CCTTCGCTGGGACCTGCTGCCCGCACAGCGACAGCTGCTTCGACGAGGTGGCCATT 120
Qy 194 ctccgctggc 251
Db 121 CTGCGCGGCG 180
Qy 252 actcat--ctacctggccaatctggagggcgaagcagcagcagcagcagcagcagc 309
Db 181 ACTCATTTCTATCTCTGGCAATCTGAGAGCGCAAGAGCAGCAGGTGGCGGCTGCT 240
Qy 310 cccggggc 365
Db 241 CCCGGGGCG 300
Qy 366 gcgcagcgtggccaa-gctggccatcatcttcagccacatgcacgcagcagcagcagc 424
Db 301 CGGACTGCTGSCCAATGCTGSCCATCATCTTCAGCCACATGCACGAGAGTGCACGC 360
Qy 425 tcttcggcgggcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 484
Db 361 TCTTCCCGCGGCG 420
Qy 485 ccttcgtgagggaaagtggcggagcccggtgtgtgctgccccctggcgtgagttt 544
Db 421 CCTTCTGAGGGAAGTTCGGAGCCCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 480
Qy 545 tcttcggcagcgtgcaacccctgtgaaaccagcgtgcacagccctggtgcgcacac 604
Db 481 TCCCTGGGACCTGCGCACCCCTGTGGAACCCAGGCTGCACAGCCCTGGCCTTGG 540
Qy 605 ttgaactcaactcagc 664
Db 541 TTGACTTCACCTGCAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Qy 665 ttcaagcctgcccacacactcctcgaagcagcagcagcagcagcagcagcagcag 724
Db 601 TTCAGCCATGCGCAACACTCCTCAAGAACTGGCAGCTCCTGGCAGTCAACCA- 659
Qy 725 acatggccttcctcactatgatg 748
Db 660 ACATGGCTTCCTACTATGATGAGG 683

RESULT 3
BE464528 522 bp mRNA EST 27-JUL-2000
LOCUS hv84hl1.x1 NCI_CGAP-Lu24 Homo sapiens cDNA clone IMAGE:3180165 3',
DEFINITION similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2.; mRNA
sequence.
ACCESSION BE464528
VERSION BE464528.1 GI:9510303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 522)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 508.
 Location/Qualifiers
 1. 522
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3180165"
 /clone_lib="NCI-CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."
 117 a 168 c 153 g 84 t

BASE COUNT 117 a 168 c 153 g 84 t
 ORIGIN
 Query Match 33.3%; Score 515.8; DB 35; Length 522;
 Best Local Similarity 99.6%; Pred. No. 2e-110;
 Matches 517; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 878 ctgcacaaacccctgtccaggtgctctgagggacagagagcggcttctactct 937
 DB 4 CCGCACAAACCCCTGTCACGGTGTCTCTGGAGGACAGAGAGCGCTTCTACTCT 63
 QY 938 accagatgaaagaccccaacccagacactgactgagctcgccaggcagacccacg 997
 DB 64 ACCAGATGGAAGAACCCACACCCAGACCTGACTGAGCTCGGCCAGGAGAACCCAGC 123
 QY 998 agcgatccacgtgtcagagagcagctgagcttactggccatggactccacttg 1057
 DB 124 AGGCATCCACGTGTACAGAGACAGCTGAGCTCTACTGGGCCATGGACTTCACATTTG 183
 QY 1058 agctctcaagatctgtgctgagagcaacagagatgaaagattgagcgtgaggcacc 1117
 DB 184 AGCTCTCAAGATCTGTGCTGAGAGAACCAAGGATGGAAGATTGAGCGGTGGGGCACC 243
 QY 1118 tgctcagactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1177
 DB 244 TGCTCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303
 QY 1178 gccgctgcagatcaagggtggtggaggccgtgagtatctaccagttccacggctcaggct 1237
 DB 304 GCCGCTCGAGATCAAGGGGTGGAGGCCGTGAGTATCTACCACTCCAGGTTCAGGCTA 363
 QY 1238 ctgctgaggaactcaggaaacagcagtgaccaggaagcaggaggttgagctggggcagg 1297
 DB 364 CTGCTGAGGACTCAGGGAACAGCAGTGACCAAGAGCGAGGTGGAGCTGGGGCAGG 423

QY 1298 tgcccttcggctctcctccattgccccccacggccagatctgccccccaggaagccagaa 1357
 DB 424 TGCCCTTTGGCTCTCTCCATTGCCCGCCAGATCTGCCCCCATGAAGCCAGAA 483
 QY 1358 atgccagccgaagtggagactcctaaaggggaactccc 1396
 DB 484 ATGCCGAGCGGAAAGTGAGACTCTTAAGGGGAAGTCCC 522
 RESULT 4
 LOCUS AI690580/c
 DEFINITION Q02all.x1 NCI-CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2207612 3', similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2.; mRNA
 ACCESSION AI690580
 VERSION AI690580
 KEYWORDS EST, GI:4901882
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 530)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 381.
 Location/Qualifiers
 1. 530
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2207612"
 /clone_lib="NCI-CGAP_Ut3"
 /tissue_type="poorly-differentiated endometrial adenocarcinoma, 2 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.45 kb. Life Technologies catalog #:
 11541-018"
 BASE COUNT 87 a 162 c 161 g 118 t 2 others
 ORIGIN
 Query Match 33.1%; Score 512; DB 12; Length 530;
 Best Local Similarity 99.6%; Pred. No. 1.5e-109;
 Matches 512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1022 agctgcagcttactggccatggactccacatttgagctctgcaagatctgctgagaga 1081
 DB 514 AGCTGCAGCTCTACTGGGCCATGGACTCCACATNTGAGCTCTGCAAGATCTGTGCTGAGA 455
 QY 1082 gcaacaagatgtgaagattgagccgtgcgggcaacctgctgagctgctgctgctgctg 1141
 DB 454 GCACAAAGGATGTGAGATTGAGCCGTGGGGCACCCTGCTCAGCTGCTGGCTG 395
 QY 1142 cctggcagactcggacagccagacctgccccctctctccgctcgagatcaaggctggg 1201
 DB 394 CTGGCAGCACTCGGACAGCCAGACCTGCCCTTCTCCGCTCGGAGATCAAGGGCTGG 335

source
 1. 398
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1991273"
 /clone_lib="NCI_CGAP_Ut3"
 /tissue_type="poorly-differentiated endometrial
 adenocarcinoma, 2 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: Salt;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.45 kb. Life technologies catalog #:
 11541-018"
 61 a 125 c 129 g 83 t

BASE COUNT
 ORIGIN

Query Match 25.6%; Score 396.4; DB 10; Length 398;
 Best Local Similarity 99.7%; Pred. No. 1.5e-82;
 Matches 397; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1132 tgcctgctcctggcagcactggacagccagacccctgccccttctgcccgtgcgagatc 1191
 Db 398 TGcctgctcctggcagcactggacagccagacccctgccccttctgcccgtgcgagatc 339
 QY 1192 aagggtggagccgtgagtatcaccagttccaggtccaggtcagctactctgaggactca 1251
 Db 338 AAGGCTGGAGGCGGTGAGTACTACCACTTCCCGGTGAGGCTACTGCTGAGGACTCA 279
 QY 1252 ggaacagcagtgaccaggaagcagggagttgagctggggcagtgccccttctcgct 1311
 Db 278 GGAACAGCAGTGCACGAGGAGCAGGGAGTTGGAGCTGGGCGAGTGCCTTTCGGCT 219
 QY 1312 cctccattgccccagcagatctgccccaggaagcccaagaatgccagccgaaa 1371
 Db 218 CTTCCATTGCCCCAGCGGCAGATCTGCCGCCCGCCAGGAGCCAGAAATCCCGAGCGGAA 159
 QY 1372 gtgagactcctaaagggaactccctccagctgcgtggcggccagccagccctgcccg 1431
 Db 158 GTGAGACTCCTAAAGGGGACCTCCCTCCAGTGCCTGGGACCCAGACCCCTGCCCGG 99
 QY 1432 gctgaagccagggccacagatgctgctcaaggggagccccaagggtggaaggggg 1491
 Db 98 GCTGAAGCCAGGGACCCAGATGCTGCTCAAGGGAGCCCAAGGSGCTGGAAGGGGG 39
 QY 1492 ttgtgaacggaataaactgcaagcctggtctgtca 1529
 Db 38 TTGTGAACGGAATAAATACTGCCAAGCCCTGGTCTGTGCA 1

RESULT 7
 AA113289/c 462 bp mRNA EST 23-DEC-1997
 LOCUS
 DEFINITION
 Zm28c07.s1 stratagene pancreas (#937208) Homo sapiens cDNA clone
 IMAGE:526956 3' similar to TR:G862411 G862411 CBL-B TRUNCATED FORM
 2. ; mRNA sequence.
 AA113289.1 GI:1664994
 AA113289
 EST
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 (bases 1 to 462)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
 B., Morris, M., Parsons, J., Prange, C., Rifkin, B., Rohlfing, J.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.,
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 9704478
 Contact: Wilson RK

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 WARNING: There is evidence that suggests that the 384-well parent
 plate of this clone contains both human and mouse derived clones.
 Thus, the origin of this clone is uncertain. This caution should be
 kept in mind should you use this clone.

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 367.

FEATURES

source

1. 462
 /organism="Homo sapiens"
 /db_xref="GDB:3918385"
 /db_xref="taxon:9606"
 /clone="IMAGE:526956"
 /clone_lib="Stratagene pancreas (#937208)"
 /lab_host="SOER cells (kanamycin resistant)"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Pancreatic adenocarcinoma cell line. Average
 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTT 3' 99 t
 BASE COUNT 76 a 141 c 137 g 99 t
 ORIGIN

Query Match 25.4%; Score 393.6; DB 1; Length 462;
 Best Local Similarity 95.3%; Pred. No. 7.1e-82;
 Matches 443; Conservative 0; Mismatches 18; Indels 4; Gaps 4;
 QY 1066 aagatctgtctgagcaacaaggatgtgaagattgagccgtgcgggacactgctctgc 1125
 Db 462 AAGATCTGTGCTGAGACCAACAAGGATGTGAAGATTGAGCGTGGCGGACCTTTCTGC 403
 QY 1126 agctgctgcctggctgctgcagcactcgacag-cacagacctgcccttctgcgcctg 1184
 Db 402 AGCTGCTGCTGGTGGCTGGCAGCCTCGACAGCCAGACCTGCCCTTCTGCCGCTG 343
 QY 1185 cgagatcaagggtgggagccgtgagttatctaccagttccacggtcaggtactgctga 1244
 Db 342 CGAGATCAAGGGGTGGGAGGCCGTGAGTATCTACCACTTCTACGGTCAAGCTACTGCTGA 283
 QY 1245 ggaactcaggaacagcagtgaccaggaagcaggaggtggagctgggcagggtgccct 1304
 Db 282 GGATCAGGGAACAGCAGTACCAGGAAGCAGGGAGTTGGAGCCTGGGACGTGCCCT 223
 QY 1305 ttggctctctccattgccccacggccagatctgccccccaggaagcccaagaaatccca 1364
 Db 222 TTNGGCTCTCTCCATTGCCCTCCACGCGCAGATCTGCCCTCCAGGAGCCCAAGAAATGCCCA 163
 QY 1365 gcgaaagttagactcctcctaagggggaactccctccagctgctgctgggacccagacc 1424
 Db 162 GCCGAAAGTGAAGTCTCTAAAGGGGAAGT-CCCNCTNTGCTGGGACCCAGGACNC 104
 QY 1425 tgccccggcctgaagccagggcaccagatgctgctcgaaggagccccaaggctgg 1484
 Db 103 NCINCCG-CTGAAGCCAGGGACCCACCATGTGCTGCTCAAGGGA-CCCAAGGGCTGG 46
 QY 1485 aaggggtgtgaaacccgaataaactgcaagcctggtctgtca 1529
 Db 45 AAGGGGTTGTGAACCCGAAATAAATGCTGCAAGCTGCTGTGCA 1

RESULT 8
 AA1463722
 LOCUS

AW463722 440 bp mRNA EST 24-FEB-2000

DEFINITION BP230013B10F1 Soares normalized bovine placenta Bos taurus cDNA
 clone BP230013B10F1 5', mRNA sequence.
 ACCESSION AW453722
 VERSION AW453722.1 GI:7033990
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 440).
 AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
 ,J.H.
 TITLE Bovine ESTs
 JOURNAL Unpublished (2000)
 COMMENT Contact: Lewin, H. A.
 W. M. Keck Center for Comparative and Functional Genomics
 University of Illinois at Urbana-Champaign
 240 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
 61801, USA
 Tel: 217 543 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu
 Funding for cattle EST sequencing was provided by the USDA National
 Research Initiative, Animal Genome Resource Project Grant AG 99-3205-8534
 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
 from Washington University Genome Center. Vector Trimming: g:
 Cross_match from Washington University Genome Center PHRAP suite.
 Sequences submitted are vector free and at least 200 bp in length.
 PCR Primers
 FORWARD: TATACGACTCACTATAGG
 BACKWARD: ATTACCTCAGTCAATG
 Insert length: 440 Std Error: 0.00
 Plate: BP230013B10 row: F column: 1
 Seq primer: AGCGGATAACAATTTCACACGGA
 High quality sequence stop: 440.
 FEATURES
 source
 1..440
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="BP230013B10F1"
 /clone_lib="Soares normalized bovine placenta"
 /sex="female"
 /lab_host="DH10B"
 /note="Orcan: Placenta; Vector: pT73pac; Site_1: EcoRI;
 Site_2: NotI: The cDNA library was contributed by the
 Soares laboratory and it was constructed and normalized
 as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806."
 BASE COUNT 100 a 142 c 120 g 78 t
 ORIGIN
 Query Match 24.0%; Score 370.8; DB 22; Length 440;
 Best Local Similarity 90.4%; Pred. No. 1.5e-76;
 Matches 396; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
 Qy 760 cgtctgcaggcctgcaggagcaagccaggcaggtacattcttcggcccgactgtactcgc 819
 Db 1 CGTCTGCAGACTTTCAGAGACAAGCCAGCGAGCTACATCTTCGSCCCAGCTGCACCTGC 60
 Qy 820 ctggggcgagtgggccatcggtatgtgagctcagatggcagcatctctcagaccctcct 879
 Db 61 CTGGGGAGTGGCCATCGGATACGTGAGCTAGCGAGCATCTTCAGACCACCTCCT 120
 Qy 880 gccacaacccctccacggtgctctggaggacagagcggtctctacctctac 939
 Db 121 CTCACAACCCCTGTTCAGGTGTCTCTGGAGGACAGAGGAGGCTTCTACCTCTAC 180
 Qy 940 ccagatggaagaccacacccagacatctgagctcgtggcgagcgagaaacccacag 999
 Db 181 CCAGATGGGAACCAACAACCCGACTCTACCGAACTCTGCCATCATGAACCGCATGAC 240

```

Query Match      23.0%; Score 356.2; DB 1; Length 435;
Best Local Similarity 94.1%; Pred. No. 3.9e-73;
Matches 381; Conservative 0; Mismatches 20; Indels 4; Gaps 1;

QY 775 agggacaagccaggcaggtacatctctccggcccgagcttactcgcctggggcagtgggcc 834
      |||
Db 1 AGGGACAAGCCAGGCGAGTACATCTTCGGCCCGAGCTTACTCGCTGGNCAGTGGGCC 60

QY 835 atcgctatgtgagctcagatggagcagctcctcagaccatccctgccaacaaacccctg 894
      |||
Db 61 ATCGNTATGTGAGCTCAGATGGCAGCATCTCGAGACCATCCCTGCCAACAACCCCTG 120

QY 895 tcccaggctcctctgaggagcagagagcagcgtctctacccctcaccagatgggaagacc 954
      |||
Db 121 TCCCAAGGTGCTCTCGAGGGAGCAGAAAGACGCGCTTACCTCTACCCAGATGGAAGACC 180

QY 955 cacaacccagacactgactgagctcggcagcagacagaccccgagcagcagcagcagcagc 1014
      |||
Db 181 CACACCCAGACCTGACTGAGCTCGGCGCAGCAGACCCCGCAGCGCATCCAGCTGTCA 240

QY 1015 gaggagcagctcagctcactgagcagcagcagcagcagcagcagcagcagcagcagcagc 1074
      |||
Db 241 GAGGAGCAGCTGCAGCTCTACTTGGGCCATGGACTCCACATTTGAGCTCTGCAAGATCTGT 300

QY 1075 gctagagcaacaagagtgatgaagattgagccgtgagcagcagcagcagcagcagcagcagc 1134
      |||
Db 301 GCTGAGACCAACAGGAGTGAAGATTGAGCCGTGCGGGCACCTGTCTCTGAGTGCTGC 360

QY 1135 ctggctgctggcagca----ctggcagcagcagcagcagcagcagcagcagcagcagcagc 1175
      |||
Db 361 CTGGCTTGCTTGGCAGAACTTGGACAGCCAGACCTTGCCCT 405

RESULT 10
AW426122      428 bp mRNA EST 07-JUL-2000
LOCUS      59653 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION      AW426122
ACCESSION      AW426122
VERSION      AW426122.1 GI:6954069
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 428)
Smith, R.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,
Bennett, G.A., Fahrrenkrug, S.C., Freking, B.A., Rohrer, G.A., Laegreid
, W.W. and Keefe, J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
V0.980904.e. Vector identified by cross_match with the -minscore 20
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCCAGTCAGCAGC
Plate: 27 row: I column: 8
Seq primer: ATTGAGTACACTATAG.
Location/Qualifiers
1. .428
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"

FEATURES
source

```

```

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT      102 a      133 c      121 g      72 t
ORIGIN

Query Match      21.7%; Score 336.2; DB 22; Length 428;
Best Local Similarity 89.4%; Pred. No. 1.8e-68;
Matches 362; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 709 gtaaacacacagggctacatggccttcctccactatgatgaggtccaagagcgtctgcag 768
      |||
Db 24 GGCAACACACCGGGTTACATGGCCITCTCTACATAGATGAGTCCAAAGCCGCTGCGAG 83

QY 769 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 828
      |||
Db 84 ACTTTCAGAGACAAGCCAGGAGCTACATCTTCGGGCCAGCTGCACTCGCTGGGGAAG 143

QY 829 tgggcatcggtctgtgagctcagatggcagcagcagcagcagcagcagcagcagcagcagcag 888
      |||
Db 144 TGGGCCATCGGATAGCTGAGCTCAGACGGCAGCATCTTGCAGACCATCCCTCTCACAAA 203

QY 889 cccctgtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 948
      |||
Db 204 CCCCTGTTCCAGTGTCTCTGGAGGACAGAGAAGAGGCTTCTACTCTACCCAGATGGG 263

QY 949 aagaccacaacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1008
      |||
Db 264 AGAACCCACACCCGAGCCTGACCGAATCTGCCACATGGAACCGCATCAGCAGATCCAC 323

QY 1009 gtgtcagaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1068
      |||
Db 324 GTGTGGAGGAGCAGCTGCAGCTCTACTGGCCATGGACTCTAGCTTTGAGCTCTGCAAG 383

QY 1069 atctgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1113
      |||
Db 384 ATCTGTCCGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 428

RESULT 11
A1758472/c    258 bp mRNA EST      16-DEC-1999
LOCUS      ty08b03.x1 NC1_CGAP_U13 Homo sapiens cDNA clone IMAGE:2278445 3',
DEFINITION      similar to contains PTR5.t3 MER22 PTR5 repetitive element ;, mRNA
sequence.
ACCESSION      A1758472
VERSION      A1758472
KEYWORDS      EST..
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 258)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1575 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers
1. .258

```



```

1 GTGAGCTCAGATGCAGCATCTCTCAGACCATCCCTCTCACAAACCTCTGTTCAGACA 60
QY 904 ctcttgaggagacagaagcgggtttctacacctaccagatggaagacccacaccca 963
Db 61 CTCTGGAAGACAAAAGGAGGCTTACCTCTACCTGTATGTTGGAAGAACCAACCCG 120
QY 964 gacctgactgagctcggcagcaggaacccagcagcagcagcagcagcagcagcag 1023
Db 121 GACCTGACGGAGCTCTCGGTAAACGGACCCCTATCAACGCATCCACGTGTGCGAGGACAG 180
QY 1024 ctgcagcttactgggcatgagctccacatttgagctctgcaagatctgtgtgagagc 1083
Db 181 CTGCAGCTGTACTGGGCCATGAACCTCACATACAGAGCTCTGCAAGATCTGTGCGGAGAGA 240
QY 1084 aacaagatggaagattgagcctgaggcagcagcagcagcagcagcagcagcagc 1141
Db 241 AACAAAGACGTGAAGATTGACCCATGCGGGGCACTGCTCTGCAGCCCGGTGCGTGGGCTA 300
QY 1142 cctggcagcactcgacagcagcagcagcagcagcagcagcagcagcagcagcagc 1185
Db 301 CCTGGCAGACTCAGACAGNCCAGACCTGCCCTTTCTGCGCCG 344

```

RESULT 14

```

AW857217 529 bp mRNA EST 19-MAY-2000
LOCUS RC2-CT0304-120200-012-e10 CT0304 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW857217
VERSION AW857217.1 GI:7952910
KEYWORDS EST.
SOURCE human.

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE

```

1 (bases 1 to 529)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

```

```

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

```

```

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

```

JOURNAL

MEDLINE

COMMENT

```

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

```

```

Tel: +55-11-2704922
Fax: +55-11-2707001

```

```

Email: asimpson@ludwig.org.br

```

```

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC2-CT0304-120)

```

```

200-012-e10&t3=2000-02-12&t4=1)

```

```

Seq primer: puc 18 forward

```

```

High quality sequence start: 17

```

```

High quality sequence stop: 507.

```

```

Location/Qualifiers

```

FEATURES

source

```

1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0304"
/dev_stage="Adult"
/Note="Organ: colon; Vector: puc18; Site: 1: Smal; Site_2:
Smal; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low

```

```

BASE COUNT 169 a 129 c 111 g 120 t
ORIGIN

```

```

Query Match 12.5%; Score 193.2; DB 25; Length 529;
Best Local Similarity 91.9%; Pred. No. 5.1e-35;
Matches 204; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

```

QY 923 acgggtttcttaccagatggaagacccacacacccagacctgactgactgactgactg 982
Db 308 ATGACTCTTAAGTGGGCCAGATGGAAGACCCACCAACCCAGACTGACTGAGCTCGGCC 367
QY 983 agcagaacccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1042
Db 368 AGGCAGAACCCACAGCAGCGCATCCACGTGTGAGAGGAGCAGCTGCTACTTGGGCCA 427
QY 1043 tggactccacatttgagctgtgcaagatctgtgtgagagcaacaaggatgtgaagattg 1102
Db 428 TGGACTCCACATTGAGCTCTGCAAAATCTGCTGTAGTACCAACAGGATGTGAGAGATTG 487
QY 1103 agcctgaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1144
Db 488 AGCGCTGCGGACACTGCTGTAGAGCTGTTCGCTAGGTGCT 529

```

RESULT 15

```

BE283475 545 bp mRNA EST 13-JUL-2000
LOCUS 601102508f1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3495025 5',
DEFINITION mRNA sequence.
ACCESSION BE283475
VERSION BE283475.1 GI:9159750
KEYWORDS EST.
SOURCE house mouse.

```

ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 545)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM8544 row: p column: 02
High quality sequence stop: 542.

```

FEATURES

source

```

1..545
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone="IMAGE:3495025"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin"
/lab_host="DH10B"
/Note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 97 a 188 c 167 g 93 t
ORIGIN

```

Query Match

```

12.1%; Score 187.8; DB 34; Length 545;

```

Best Local Similarity 70.2%; Pred. No. 9.4e-34;			
Matches 384; Conservative 0; Mismatches 142; Indels 21; Gaps 9;			
QY	74	cagtcaggatgctgagcctcctagagagcaatgctgaccccgctgtccgtgagtc	133
Db	1	CGGTGAAGTGTCTCCAGGCCCTAGAGAGCAATGAGGATCCCAAGATGTCACGGGC	60
QY	134	cccttcgctgctgggacctgctgccccgcacagcgagctgcttcgagaggtggccatt	193
Db	61	CCCCGTCCCTGGGGACCTGCTGCCCGCACCGCGAGCTACTTGGAGAGGTGGCAAGG	120
QY	194	ctcggcgagcgccgagcgagcgcccgggggtcccggggctctctgggactttctac	253
Db	121	CCGGCGCGAGGCCACAGGAA---GACCCGAGGGTCCCGGTGGCGCCGATGACTTTCGG	177
QY	254	tcattacctggccaatctggaggcccaagagcagggaggtggc-cgcgctgctgctccc	312
Db	178	CCATCTACCTGGCCATCTGGAGGTCAAGGCGAGCGAGCTGGCGAGCTCCTGCCACCC	237
QY	313	cgggggcoga-aggagtgccaaagacagctcttcocggcgggctccagact--caggcga	369
Db	238	CGAGGCAAAACCCAGGACGTGAACCCAGATGTTTCCGGGAGGGCTCCAGACTTCACGGGA	297
QY	370	cagctggccaa-gctggccatcatcttcagc---cacatgcacgagagctgcagcact	425
Db	298	CAACTGGCCAAACGCTGGCCCTCATCTTCAGCTCACATGCACGGCGAGCTTGAGCGCACT	357
QY	426	cttccccg-gggcaagtactgtggacacatgtaccagctcaccaggccccccgccaca	484
Db	358	CTTCCCTGCTGGGAAGTACCGTGGCACCTGTACCAGCTACCAAGGGCTCTGCCACA	417
QY	485	ccttcggagggaaagtgcggagcc-----cggtgtgctgctccctgggctgagtt	536
Db	418	TTCTCTGGCAGCGCAGAAATTGTGGAGTTCGGGTGCGCTGCTTCCCTGGCGCTGCAGTTCC	477
QY	537	tgagtcctcctctggggcacctgccaccctg-tggaaaccaggtgcacagccctggccttgc	595
Db	478	AGCTCCCTCCTGTGCTGCTGCTGCCACCTGCTGTGAACCCAGGCCCCACCATGCAGGCCTTGC	537
QY	596	gcaccac	602
Db	538	GGTCCAC	544

Search completed: December 7, 2000, 12:41:38
Job time: 24581 sec

Blank

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 7, 2000, 06:01:08 ; Search time 3361.48 Seconds
(without alignments)
2010.021 Million cell updates/sec

Title: US-09-434-708-1
Perfect score: 1547
Sequence: 1 cggagcgtcccatggctct.....caaaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1033670 seqs, 2183789903 residues 2067340
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_bai.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pli.*
8: gb_pl2.*
9: gb_pri.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sy.*
14: gb_un.*
15: em_fun.*
16: em_hum1.*
17: em_hum2.*
18: em_in.*
19: em_om.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: gb_htg1.*
31: gb_htg2.*
32: gb_in1.*
33: gb_in2.*
34: em_ba1.*
35: em_ba2.*
36: em_hum3.*
37: em_hum4.*
38: gb_pr4.*
39: gb_htg3.*
40: gb_htg4.*
41: gb_htg5.*
42: gb_htg6.*
43: gb_htg7.*

F22931

44: em_htg1.*
45: em_htg2.*
46: em_htg3.*
47: em_hum5.*
48: gb_pl3.*
49: gb_pr5.*
50: gb_htg8.*
51: gb_htg9.*
52: gb_htg10.*
53: gb_htg11.*
54: gb_htg12.*
55: gb_htg13.*
56: gb_htg14.*
57: gb_in3.*
58: gb_htg15.*
59: gb_htg16.*
60: gb_htg17.*
61: em_htg4.*
62: em_htg5.*
63: em_htg6.*
64: em_htg7.*
65: em_hum6.*
66: gb_htg18.*
67: gb_htg19.*
68: gb_htg20.*
69: gb_htg21.*
70: gb_htg22.*
71: gb_htg23.*
72: gb_vl1.*
73: gb_vl2.*
74: gb_ba3.*
75: em_htg8.*
76: em_htg9.*
77: em_htg10.*
78: em_htg11.*
79: em_htg12.*
80: em_htg13.*
81: em_htg14.*
82: em_htg15.*
83: em_htg16.*
84: em_htg17.*
85: em_htg18.*
86: em_htg19.*
87: em_htg20.*
88: em_htg21.*
89: em_htg22.*
90: em_htg23.*
91: gb_pr6.*
92: gb_pr7.*
93: gb_sts1.*
94: gb_sts2.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
1	1539.6	99.5	1575	9	AB028645		AB028645 Homo sapi
2	1525.2	98.5	1584	11	AF117646		AF117646 Homo sapi
3	1239.2	80.1	1446	11	AF117647		AF117647 Homo sapi
4	340	22.0	3090	91	HSCCBL		X57110 Human mRNA
5	329.2	21.3	2661	57	DMU87925		U87925 Drosophila
6	327.6	21.2	2529	57	DMAJ3175		AJ223175 Drosophila
7	323.6	20.9	3241	91	HSU26711		U26711 Human cbl-b
8	323.6	20.9	3354	91	HSU26712		U26712 Human cbl-b
9	323.6	20.9	3982	91	HSU26710		U26710 Human cbl-b
10	270.6	17.5	2808	12	MMCCBL		X57111 Mouse mRNA
11	248.2	16.0	2711	72	CASNS1		J04169 Mouse Cas N
12	221.4	14.3	237653	60	AC073760		AC073760 Mus muscu

Db 810 AGCCAGGAGTACATCTTCGGCCAGCTGATCGCCTGGGCGACGTGGCCATCGGC 869
QY tatgtgagctcagatggagcattcctgagaccatccctgccaacaaacccctgtccag 900
Db 870 TATGTGAGCTCAGATGGCAGCATCTGAGACCATCCCTGCCAACAACCCCTGTCCAG 929
QY gtgctcctgagagcagagacagcgtcttacccttaccctgagagagagagagagagag 960
Db 930 GTGCTCCTGAGGAGCAG 989
QY ccagacctgactgagctggcagcagagagagagagagagagagagagagagagagagag 1020
Db 990 CCAGACCTGACTGAGCTGGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
QY 1021 cagctcagcttactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1080
Db 1050 CAGCTCAGCTTACTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1109
QY 1081 agcaacaagatgtgaagattgagcgtggcagcagcagcagcagcagcagcagcagcagc 1140
Db 1110 AGCAACAAGATGTGAAGATTGAGCGCTGGGCGACCTCTCTGAGCTGCTGCCGTGGCT 1169
QY 1141 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200
Db 1170 GCTGCGACACTCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1229
QY 1201 gagcgcgtgagctatccagcttccagcgtccagcgtccagcgtccagcgtccagcgtcc 1260
Db 1230 GAGCGCGTGAGTATCTACCATCTACCATCTACCATCTACCATCTACCATCTACCATCT 1289
QY 1261 agtgaccaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320
Db 1290 AGTGACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1349
QY 1321 ccccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380
Db 1350 CCCCACGCGCAGATCTGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409
QY 1381 ctacagggagacccctccagctgctgagcagcagcagcagcagcagcagcagcagcagcagc 1440
Db 1410 CTAAAGGGGAACCTCCCTCCAGCTGGCTGGGACCCAGAGAGAGAGAGAGAGAGAGAGAG 1469
QY 1441 ccagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500
Db 1470 CCAGGCGACCCAGATGTGCTCAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1529
QY 1501 cgaataaactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1546
Db 1530 CGAATAAAGTCCAGAGCTGGTCTGTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1575

RESULT 2
AF117646
LOCUS AF117646 1584 bp mRNA PRI 17-JUN-1999
DEFINITION Homo sapiens long CBL-3 protein (CBL-3) mRNA, alternatively spliced isoform, complete cds.
ACCESSION AF117646
VERSION AF117646.1 GI:4959420
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1584)
Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M., Penninger,J., and Lipkowitz,S.
cbl-3: a new mammalian cbl family protein
Oncogene 18 (22), 3365-3375 (1999)
928203
MEDLINE
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 1584)
Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M., Penninger,J., and Lipkowitz,S.

Direct Submission
Submitted (04-JAN-1999) Genetics Dept., Medicine Branch, DCS, NCI,
Bethesda: Naval Hospital, Bethesda, MD 20889, USA
FEATURES
source
1..1584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2"
/cell_line="CF-PAC1"
/tissue_type="pancreatic adenocarcinoma"
gene
1..1584
/gene="CBL-3"
64..1488
/gene="CBL-3"
/note="long form, alternatively spliced; cbl family protein; structurally related to the Sli-1 protein, encoded by the cbl gene of *Caenorhabditis elegans*"
/codon_start=1
/product="long CBL-3 protein"
/protein_id="AAD34341.1"
/db_xref="GI:4959421"
/translation="MALAVPWGWEAREALGRVLMQLRLEEQCDRLSVSPSL
RDLPTAOLLREVAHRAAGGGGPGGSGDFLLIYLANLEAKSRQVAALLPPRG
RSANDELFRAGSLRLROLAKLAIIFSHMAELHAFPGKYCGHMYQTKAPATFW
RESCGARCVLPWAEFESLLGTCHPVEGCTALALATTIDLTCSGHVSIIEFDVETRLF
QNPMTLLKNWQLLAVNHPGMAFLTYDEVQRLOACDRDPSYIFRPSRGLQWAIG
YVSSDSILQTIIPANKPLSGVLEGGKDDGYFLDPDKTHNPDLTGLQAEPPQRIHVS
EQQLQYWMADSTFEKCAESNKDKIEPCGHLKSCCLAAWHSQDTQPFCKE
IKGWEAVSYFYQATAEKSDGNSDQGELELQVPLSAPLPFRPLPPRPKPRNA
OPKVRLLKGNSPPAALGPDPAPA"

BASE COUNT 303 a 538 c 482 g 261 t
ORIGIN
Query Match 98.6%; Score 1525.2; DB 11; Length 1584;
Best Local Similarity 99.8%; Pred. No. 4.5e-272;
Matches 1527; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 cgcagggctccatcgctctgagcgtggccgctggggggcagcagtgaggagagagccgcg 60
Db 52 CGCGAGGCTCCATGGCTCTCGCGGTGCCCGCTGGGGGCGACAGTGGAGAGGCGCGC 111
QY 61 gccctggcgccggcagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 120
Db 112 GCCCTGGCGCGGCGAGTCAGGATGCTGCGAGCGCTAGAGAGCAATGCTGCGACCCCGG 171
QY 121 ctgtccgtgagtcgcccttcgctgagcagcagcagcagcagcagcagcagcagcagcagcagc 180
Db 172 CTGTCCGTGAGTCCCGCTTCGCTCGGGACCTGCTGCCCGCGACAGCGCAGCTGCTTGA 231
QY 181 gaggtggccattctcggcgccgagcagcagcagcagcagcagcagcagcagcagcagcagc 240
Db 232 GAGGTGGCCCATTCCTCGCGGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 291
QY 241 ggggactttctactctactctactctgagcagcagcagcagcagcagcagcagcagcagcagc 300
Db 292 GGGGACTTTCTACTCATCTACCTGGCCAACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 351
QY 301 ctgctgctctcccgggcgagagagtgcccaacagcagcagcagcagcagcagcagcagcagc 360
Db 352 CTGCTGCTCTCCCGGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 411
QY 361 ctccagcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
Db 412 CTCAGGCGACAGCTGGCCAACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 471
QY 421 gcaactctcccgggcgagagagtgctgagcagcagcagcagcagcagcagcagcagcagcagc 480
Db 472 GCACCTCTTCCCGGGGAGGAGTACTGTGGACACATGTACCACTGTACCAAGGAGGAGGAGGAG 531
QY 481 cacacctctcagggagagagtgccgagcagcagcagcagcagcagcagcagcagcagcagcagc 540

Db 532 CACACCTTTCTGGAGGAAAGTTGGGAGCCCGGTGTGTGTCTGCCCTGGGTGAGTTTGG 591
 Qy 541 tccctctggcagcactgcacccctgtggaaccagctgcacagccctggcctgcgcacc 600
 Db 592 TCCCTCTGGGCACTGCCACCCCTGTGGAACCAAGCTGCACAGCCCTGGCTTGGGCACC 651
 Qy 601 accattgacctcaactgcaggggcaagtgccatcttcagttcgagttcgagctcttcaccag 660
 Db 652 ACCATTGACCTCAGCTGCAGCGGCACGTGTCCATCTTCGAGTTCGAGCTTCACACAGG 711
 Qy 661 ctcttcagccatggcaacactcctaagaactggagctcctggagtcagtcacaccacca 720
 Db 712 CTCCTTCAGCCATGGCCAACTCCTCAAGAACTGGCAGCTCTGGCAGTCAACACCCA 771
 Qy 721 ggcacatggccttcctcaccctatgatgaggtccaaagagcgtctgcagggcctgcaggag 780
 Db 772 GGTACATGGCTTCCTCACCTATGATGAGTCCAGAGCGTCTGCAGGCTGCAGGGAC 831
 Qy 781 aagccaggcagttacattctccggccagctgtactcgtctggggcagtggtgcacatggc 840
 Db 832 AAGCCAGGCACTTACATCTTCGGCCAGCTGTACTCGCTGGGCGAGTGGGCATCGGC 891
 Qy 841 tatgtgactcagatggcagcactcctcagaccatccctcccaacaaacccctgtccag 900
 Db 892 TATGTGAGCTCAGATGCGCATCTCTGCAGACCATCCCTGCCCAACACCCCTGTCCAG 951
 Qy 901 gtgctctggaggagacagagagcgtcttaoctctaccagatggaagaccaccaac 960
 Db 952 GTGCTCTGGAGGACAGAGAGCGGCTTACTCTACCCAGATGGAAGACCCCAAC 1011
 Qy 961 ccagacctgactagctcgccaggcagagacccagcagcgtccacatgctgcagagag 1020
 Db 1012 CCAGACTGACTGAGCTGGCCAGGAGACCCAGAGCGCATCCAGTGTCCAGAGGAG 1071
 Qy 1021 cagctgcagcttactgggacatggactccacatttgagctctgcagatctgtgtgag 1080
 Db 1072 CAGTGCAGCTACTGGGCACTGGAGTCCACATTTGAGCTGTCAAGATCTGTGTGAG 1131
 Qy 1081 agcaacagatgtgaagatgagcgtgagggcagcactgctctgcagctgctgcctggt 1140
 Db 1132 AGCAACAGATGTGAAGATTGAGCTGGCGGACCTGCTCTCAAGCTGCTGCTGGCT 1191
 Qy 1141 gcttggcagcactcggagacccagacactgccccttctgctgctgcagatcgaaggtg 1200
 Db 1192 GCCTGGCAGCCTCGGACACCCAGACTGCGCCCTTCTGCGGCTGCAGATCAAGGCTGG 1251
 Qy 1201 gaggccctgagtatctacagttccagcgttcaggtactgctgagactcagggaaacagc 1260
 Db 1252 GAGCCGCTGAGTATCTACAGTTCTACGCTCAGGCTACTGCTGAGACTCAGGGAACAGC 1311
 Qy 1261 agtgaccaggaagcagggagtgagctggcagtgcccttctgctgctctccatg 1320
 Db 1312 AGTGACCAAGAGCAGGAGATTGAGCTGGGCGAGGTCGCCCTTTTCGCTCTCCATG 1371
 Qy 1321 ccccccagccagatctgccccccaggaagccagaaatccagccagccgaagtgcagactc 1380
 Db 1372 CCCCCAGGCGCAGATCTGCCGCCCCAGAGAACCCAGAAATGCCAGCCGAAATGTGAGACTC 1431
 Qy 1381 ctaaaaggggaactccctccagctgcgctggagcccccagagacccctcccccgcctgaag 1440
 Db 1432 CTAAAGGGGAACCTCCCTCCAGTGGCTGGGACCCCGAGACCCCTGCCCGGCTGAAGC 1491
 Qy 1441 ccaggggcaccagatgtgctcgaaggagccccaagggctggaaggggtgtgaaac 1500
 Db 1492 CCAGGGCAACCCAGATGTGCTCAAGGGAGGCCCAAGGCTGGAAGGGGTTGTGAAC 1551
 Qy 1501 cgaataaactgccaagcctgtgtctca 1530
 Db 1552 CGAAATAAACTGCCAAGCCCTGTCTGTCCA 1581

LOCUS AF117647 1446 bp mRNA PRI 17-JUN-1999
 DEFINITION Homo sapiens short CBL-3 protein (CBL-3) mRNA, alternatively spliced isoform, complete cds.
 ACCESSION AF117647
 VERSION AF117647.1 GI:4959422
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1446)
 AUTHORS Keane.M.M., Ettenberg.S.A., Nau.M.M., Banerjee.P., Cuello.M., Penninger.J. and Lipkowitz.S.
 TITLE cbl-3: a new mammalian cbl family protein
 JOURNAL Oncogene 18 (22), 3365-3375 (1999)
 MEDLINE 99289203
 REFERENCE 2 (bases 1 to 1446)
 AUTHORS Keane.M.M., Ettenberg.S.A., Nau.M.M., Banerjee.P., Cuello.M., Penninger.J. and Lipkowitz.S.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-1999) Genetics Dept, Medicine Branch, DCS, NCI, Bethesda Naval Hospital, Bethesda, MD 20889, USA
 FEATURES
 source Location/Qualifiers
 1..1446
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19q13.2"
 /cell_line="CF-PAC1"
 /tissue_type="pancreatic adenocarcinoma"
 1..1446
 /gene="CBL-3"
 64..1350
 /gene="CBL-3"
 /note="short form, alternatively spliced; cbl family protein; structurally related to the Sli-1 protein, encoded by the cbl gene of *Caenorhabditis elegans*"
 /codon_start=1
 /product="short CBL-3 protein"
 /protein_id="AAD34342.1"
 /db_xref="GI:4959423"

gene

CDS

BASE COUNT 276 a 493 c 443 g 234 t
 ORIGIN
 1 cgcagagctccatggctctggtggccctggggggcagcagtggaagggccgcg 60
 52 CGCAGAGCTCCCATGGCTCTGGCGGTGGCCCGTGGGGCGACAGTGGGAGGCCCCG 111
 61 gccctggcggggcagtcaggatctgcagcgcctagaagcaatgcgtcgaccgcccg 120
 112 GCCCTGGCGGGCAGTCAGATGCTGCAGCGCTAGAGCAATGCGTCGACCCCCGG 171
 121 ctgctcctgagctccctctgctgctggggaactgctgcccgcacagcagctctaga 180
 172 CTGTCCCTGATGCCCTTCTGCTCGGGACCTGCTGCCCGGCACAGCGAGCTGCTCGA 231
 181 gagtggtgcccattctcgccggggcgccgaggggggggggggggggggggggggggg 240
 232 GAGTGGCCCATTCCTCGCGGGGGCGGGCGGAGGGGGGGGGGGGGGGGGGGGGGGGGT 291

Query Match 80.1%; Score 1239.2; DB 11; Length 1446;
 Best Local Similarity 90.8%; Pred. No. 2.8e-219;
 Matches 1389; Conservative 0; Mismatches 3; Indels 138; Gaps 1;

QY	241	ggggacttctactcatctaccctggccaatcttgaggccaaagacagcaggtggccgagcgcg	300
Db	292	GGGGACTTTCCTACTCATCTACTCTGGCAATCTTGAGAGCCCAAGACGACGAGTGGCCGCG	351
QY	301	ctgctgcctcccgggcccaagagtgcaacgacgagctcttcggcgcggtcccaga	360
Db	352	CTGCTGCCCTCCCGGGCCGAGGAGTGCACACGACGAGCTCTTCGGGGGGGCTCCAGA	411
QY	361	ctcaggcgcagcctgggccaagtggccatcatcttcagccacatgcacgcgcagagctgcac	420
Db	412	CTCAGGCGACAGCTGGCCAAAGCTGGCCATCATCTTCAGCCACATGCACGACAGTGCAC	471
QY	421	gcaactctcccggggcaaaagtactgtgacacatgtaccagctcaccaaagcccccgc	480
Db	472	GCACCTCTCCCGGGGAAAGTACTGTGACACATGTACAGCTCACCAGGCCCCGCC	531
QY	481	cacacctctgagggaaagttagcgagccggtgtgtctgacctgggctgagtttag	540
Db	532	CACACCTTCGAGGGAAAGTTGGGAGCCGGTGTGTGTCGCTTGGCTGAGTTTGAG	591
QY	541	tcctctctggcaactggcaacctgtgaacacaggtgcacagccctgccttgcgcacc	600
Db	592	TCCCTCTGGGCACCTGCCACCCCTGTGGAACCAAGCTTGACACCTTGCCCTTGGCGACC	651
QY	601	acattgacctcacctgcagcgggcagctgtccatctcttgagttgcagctcttcaccag	660
Db	652	ACCAITGACCTCACCTGCAGCGGCGTGTCCATCTTTCAGTTTCAGCTCTTCACCAAG	711
QY	661	ctcttcagccactgcccacactcctcaagaactggcagctctgcgcagtcacacacca	720
Db	712	CTCTTTACGCCATGGCCACACACTCCTCAAGAACTTGGCAGCTCTGGCAGTCAACACCCA	771
QY	721	ggctacatggcctctctcacctatgatgaggtccaaagcgtctgcagcctgcagggac	780
Db	772	GGCTACATGGCCTTCCTCACCTATGATGAGGTCCACAGAGCTGTGCAGGCGCTGCAGGAC	831
QY	781	aagccgggcagttacatcttcggcccagctgtgtactgcctcgtgggcagtgggccatcggc	840
Db	832	AAGCCAGGCAG-----	842
QY	841	taigtgagctcagatggcagcactcgcagaccctcctgccacaaacccctgtcccg	900
Db	843	-----	842
QY	901	gtgctcctggaggacagaaggcggctctacctctaccagatggaagaccacac	960
Db	843	-----CTACCTCTACCCAGATGAAAGACCCACAAC	873
QY	961	ccagactgaactgagctcggccaggacagaacccagcagcgatccacgtgtcagaggag	1020
Db	874	CCAGACCTGACTGAGTCTGGCCGCGACGAACCCACGACGCATCCACGTGTGAGAGGAG	933
QY	1021	cagctcagctctactgggccatggactccacatttgactctbgaagatctgtgctgag	1080
Db	934	CAGCTGCAGCTCTACTGGGCCATGGACTCCCAATTTAGCTTCTGCAAGATCTGTGCTGAG	993
QY	1081	agcaacaaggatgtgaagattgagccgtgcgggcaactgtctcgcagctcctcctggct	1140
Db	994	AGCAACAAGGATGTGAAGATTGAGCCGTGCGGGCACCTGCTCTGCACCTGCTGCTGCTG	1053
QY	1141	gcctggcagcactcggacagccagacctgcacctctgcgcctgcgcagatcaaggctg	1200
Db	1054	GCCTGGCAGCACTCGACAGCCAGACCTGCCCTTCTGCCCTTCGCGAGATCAAGAGGCTG	1113
QY	1201	gagccctgagtatctaccagttccaggtcaggctactgtctgaggactcagggaacagc	1260
Db	1114	GAGCCCTGAGTATCATCCAGTCTACGGCTCAGGCTACTGCTGAGGACTCAGGGAACAGC	1173
QY	1261	agtgaccaggaagcgaggagttagctggggcgaggtgcacctcttcgctctccatg	1320
Db	1174	AGTACACAGGAAGCGAGGAGTTGGAGCTGGGGCAGGTGCCCTTTTCGGCTCTCCTCATG	1233
QY	1321	ccccacgcccagatctgccccccaggaagccagaaatgccagccgaaagtgaagactc	1380

DB	1234	CCCCACGCCAGATCTGCCCCCGCAGGAAGCCCGAAGATGCCAGCGAAGTGAGACTC	1293
QY	1381	ctaaaggggaactcccctccagctgcgtgtaggacccagagaccctgcccgccctgaagg	1440
DB	1294	CTAAGGGGAATCCCTCCAGCTCGCTGGAGACCCAGGACCTGCCCGGCTGAAGG	1353
QY	1441	ccagggcaccagatgtgtctcaaggaggagccccaagggttggaagggggttgaaac	1500
DB	1354	CRAGGCACCCAGAGTGTCTGCTCAAGGAGCCCGCCCAAGGCTGGAAGGGGTTTGAAAC	1413
QY	1501	cgaataaaactgccaaagcctgtctgtctcaa	1530
DB	1414	CGAATAAACTGCCAAGCCTGGTCTGTCTCA	1443
RESULT	4		
HSCBCL		3090 bp	mRNA
LOCUS			PRI
DEFINITION			12-DEC-1992
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
FEATURES			
source			
mRNA			
gene			
CDS			

```
/db_xref="Swiss-Prot:P22681"
/trans_xref="MAGNUS66AGGTGGSGGSLGLMKDAQPHHHHHHLS
PHPGTVDKMVKCKMLDKVLCNPKLANSPPYIDLPLDYOHURLTILSR
EGMNETUGENEYFVWENLKKTKQTISLTKKEGKERMYESNPRLNLTSLF
MLAEKGLFPFSGQGTFTIRKADAEFWKAGKEKTLVPWKGFRQALHEVHPTSSG
LEAMALKSTIDLTCDNYISVEFDITRLEQPSLNNWNSLAVTHPEGYNAFTIDE
```


/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3"
/dev_stage="embryonic"
/map="662"
/clone="pbluescript-1255"
328..1674
/gene="cbl"
328..1674
/gene="cbl"
/codon_start=1
/protein_id="CA11149.1"
/db_xref="GI:2739273"
/translation="MATRGSRTVRSQPKIPFSLSLKLGHAISEACVSRSLSTDKKTL
EKTMLKDVVKLQOQKRNKLNKSPFILDILPDIYQRLIYKSNEDQMLHANEH
FNFIINLRKQKAIKFKEGKMFENSHYRNLTSLVFSHMLSELAIFPNP
VFAGDQFRITKADAADWKSNGSNLFWPKIFROELNWHPIIISLEAMAKITDL
TNDFTSNREDFVTEFLFOPWTLRNWOLAVTHPQVAFILYDSVAKARLYILKA
GSTVFRSLSTRGQNAIGVTAEGEILQIPONKSLCOALLDGHREGFVLYPDQAYN
PLUSSAVQSTEDHITVITQELICEMSTFQCKICAENDKDIRIECGHLLCTPC
LTSQVDSQSGQPCFRAEIKGTQIVDAFDFRKHNRNVTNGROOQGEEDIEV"
BASE COUNT 657 a 561 c 564 g 647 t
ORIGIN

Query Match 21.28; Score 327.6; DB 57; Length 2529;
Best Local Similarity 51.68; Pred. No. 3.9e-51;
Matches 541; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

Qy 328 gccaacagcagctctccgggggggtccagactccaggggagcagctggccaaagctggcc 387
Db 697 GGCAGGAGAGAGATGTCGACGAGAACTCCCACTACCGCGCAATCTACCAAGCTCAGC 756
Qy 388 atcatcttgcacatgcacgcagctgcacgactcttcccccgggggaaagtctgt 447
Db 757 CTGGTCTTCTCCACATGCTCAGCGAACTGAAGGCCATATTCGCCAACGGTGTCTTTGCC 816
Qy 448 gacacatgtaccagctcaccagggcccccgcacacctcttgaggagaaagtggga 507
Db 817 GGGATCAATTTGGATACACCAAGCGGATCGCGCTGACCTTTGGAGAGACAACTTCGGT 876
Qy 508 gccgggtgtgtgcctgggtggctgagttgagttccctcctgggcaactggccacctgtg 567
Db 877 AACAGCATATGGTTCCTCGAAGATCTTCGGCGAGGAGCTTAACAAAGTACATCCGATA 936
Qy 568 gaaccagctgcacagctggctggcttgcgaaccacattgacacctcagcgggcag 627
Db 937 ATCTCGGCGCTGGAGGCCATGGCCCTAAAGACCACTATCGATCTTACCTGCAAGACTTC 996
Qy 628 gtgtccatcttcgagttcagctgtttccagggctctttcagccatggcccaactctc 687
Db 997 ATTTCACACTTCGAGTTCGACGTCTTCACACGCTTTCACGCTTGGGTGACCGCTGCTA 1056
Qy 688 aagaactggcagctcctggcagtcacacacccagcagctacatggccttccactatgat 747
Db 1057 CGCAACTGCGAGATTCTGGCGCTCACATCCGGGCTACGTGGGGTTCCTCACATACGAC 1116
Qy 748 gaggtcccaagagcgtctcagggcctggcagggagacagcagcagcttaccatctccggccc 807
Db 1117 GAGGTGAGGTTGGGCTACAGCGCTACATCTCAAGGGGGGCGAGTACGTTTCCGGGCTC 1176
Qy 808 agcttactcgtcctggggcagtgggccatcggtctatgtgagtcagatggcagcagctgt 867
Db 1177 TCCTGACGCGATTTGGGCAATGGGCCATCGGCTACGTAAGTTCGCGAGGAGAGATTCTG 1236
Qy 868 cagaccatcctggcacaacacccctgtccaggtgtccctggaggagacagagcagcgc 927
Db 1237 CAGACAAATCCCTCAGAACAAATGCGCTGTGGCGGCGCTGCTCGATGGCCATCGCAGAGGC 1296
Qy 928 ttctactctaccagatggaaagaccacacccagacacctgagctgagcggcagca 987
Db 1297 TTCTACTTGTACCCAGATGGCCAGGATACATCCGGATCTGTCTGCTGCTCCGCTTCAAAGT 1356

Qy 568 gaaccagctgcacagcctgctgctggcaccacattgacacctcagctcagcgggcag 627
Db 923 ATCTCGGCGCTGGAGCCATGGCCCTTAAGACCACTATCGATCTTACTGTCAAGCACTTC 982
Qy 628 gtgtccatcttcgagttcagctgtttccagggctctttcagccatggcccaactctc 687
Db 983 ATTTCACACTTCGAGTTCGACGTCTTCACACGCTTTCACGCGCTTGGGTGACACTGCTA 1042
Qy 688 aagaactggcagctcctggcagtcacacacccagcagctacatggccttccactatgat 747
Db 1043 CGCAACTGCGAGATTCTGGCGCTCACATCGGGCTAGTGGCGTTCCTCACATACGAC 1102
Qy 748 gaggtcccaagagcgtctcagggcctgcagggagacagcagcagttacatcttccggccc 807
Db 1103 GAGGTGAAGGCTCGCTACAGCGCTACATCTCAAGGGGGGCGAGTACGTTTCCGGGCTC 1162
Qy 808 agctgtactcgcctggggcagtgggccatcgtgctatgtgagctcagatggcagcactctg 867
Db 1163 TCTGTACGCGGATTTGGGCAATGGGCCATCGGCTACGTAACTGCCGAGGAGAGATTCTG 1222
Qy 868 cagaccatccctgcacacacccctgtccaggtgctcctggagggagacagagcagcgc 927
Db 1223 CAGACAAATCCCTCAGAACAGTGGCTGTGTCGAGGGCTGCTCGATGGCCATCGAGAGGC 1282
Qy 928 ttctactctaccagatggaaagaccacacccacagcagcagctgagctggccagcagca 987
Db 1283 TTCTACTTGTACCCAGATGGCCAGGCTTACATCCGGATCTGTCTGCTCGCTTCAAAGT 1342
Qy 988 gaaccacagcagcgcacatgcagtgagagcagcagctcagctctactctggggccatgac 1047
Db 1343 CCACAGAGGACACATACCGTTTCCCAAGAGCAATACGAACTATACGTGAATGGC 1402
Qy 1048 tccacattgagctgcagatctgtgtgagagcaacagagatggaagttaggcg 1107
Db 1403 AGCACCTTTTCAGTGTGCAAAATTTGGGGAGAGACAGACAAAGATATCGGCATCGAGCC 1462
Qy 1108 tggggcacctgtcagctcctcctggctgctgca--gcactcgacagcag 1164
Db 1463 TGTGGCCACTTGTGTGCACTCCCTGCTTACCTCTGGCAAGTGTGATTCGAGGAGCAG 1522
Qy 1165 acctgccccttgcgctgcgagatcaagggtggga 1202
Db 1523 GGCTGCCCTTCTGTCGGGCGCAATCAGGGGACCGA 1560

RESULT 6
LOCUS DMAJ3175 2529 bp mRNA INV 28-DEC-1997
DEFINITION Drosophila melanogaster mRNA for cbl gene.
ACCESSION AJ223175
VERSION AJ223175.1 GI:2739272
KEYWORDS cbl gene; proto-oncogene.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
1 (bases 1 to 2529)
Hime,G.R.
Direct Submission
Submitted (22-DEC-1997) Hime G.R., Trescowthick Research
Laboratories, Peter MacCallum Cancer Institute, Locked Bag 1, A
Beckett St., Melbourne VIC, 3000 AUSTRALIA
2 (bases 1 to 2529)
Hime,G.R., Dhungat,M.P., Ng,A. and Bortell,D.D.
D-Cbl, the Drosophila homologue of the c-Cbl proto-oncogene,
interacts with the Drosophila EGF receptor in vivo, despite lacking
C-terminal adaptor binding sites
Oncogene 14 (22), 2709-2719 (1997)
JOURNAL 9732115 Location/Qualifiers
MEDLINE 1. .2529
FEATURES source


```

QY 1263 tgaccaggaaggcaggaggttg 1284
|||||
Db 1639 TGACCCCTTTGGCATGCCGATG 1660

RESULT 8
HSU26712 3354 bp mRNA PRI 30-SEP-1996
LOCUS Human cbl-b truncated form 2 lacking leucine zipper mRNA, complete
DEFINITION cds.
ACCESSION U26712
VERSION U26712.1 GI:862410
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3354)
AUTHORS Keane,M.M., Rivero-Lezcano,O.M., Mitchell,J.A., Robbins,K.C. and
Lipkowitz,S.
Cloning and characterization of cbl-b: a SH3 binding protein with
homology to the c-cbl proto-oncogene
Oncogene 10 (12), 2367-2377 (1995)
MEDLINE 95303504
REFERENCE 2 (bases 1 to 3354)
AUTHORS Lipkowitz,S., Keane,M.M. and Mitchell,J.A.
Direct Submission
TITLE Submitted (10-MAY-1995) Stan Lipkowitz, Navy Medical Oncology
Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval
Hospital, Bethesda, MD 20889, USA
FEATURES
source
1..3354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="breast cancer cell line ZR75-1"
/map="3q"
/chromosome="3"
323..2635
/note="similar to c-cbl proto-oncogene product, Swiss-Prot
Accession Number P22681"
/codon_start=1
/product="cbl-b truncated form 2"
/protein_id="AAB09293.1"
/db_xref="GI:862411"
/translation="MANSWNGNPGGNGNPKRGRILGIIDAQDVGPPKQAAADR
TVEKWLMDKVVRCNPKLPILDPDYQHLRLILSLDNDKLAQLS
ENEYKIVIDSLMKSKRAIRLFKEGKERMVEEQSDRRNLKLSLIFSHLAKAI
FNGQPGQDNFRITADAAEFKRFKFGDKTIAPKVFQCLHEVHQLSSSLEAKALK
TIDLCDIISVFEDIDTRLPQPGWSILRNWNLAVHPGMAFLIYDEVKARLQY
STKPSYIFRLSCTRLGOWAIGVYTGDNILLOTIPHNAPLFCALIDGREGFYILPDG
RSYNPDLTGLCPTPHDHKVTQEQELYCEMGSTFOLCKICAENDKDKVTEPCGHLM
CISLTAWQESDQGFPCFCEIKGTEPIIVDPFDRDEGRCCSIIIDFGMPMLDLD
DDDDREESLMNRLANVRKCTDRONSPTSPGSSPLAORRKPQDPLOIPLSLPPVP
PLDLIOKGIIVSPGSPGSPKSPCMVRKODKPLPAPPPLRPPPPPPPPPPPI
PNRLSRIHHVSPSPDPKPLPAPKPLPAPKPLPAPKPLPAPKPLPAPKPLPAPK
VNGRHSVGSQVPLMKRHRHDLPLEGAKVPSNGLGSEEDVPRLSFPFPPVITLLP
SIKCTGPLANSLSKTRIDPVEEDDEIKIPSHSPVLSNSQPSHCHNAPPVPRSCDNGH
CMLNGTHGFSSEKKSNIPDLISYLXGTIRI"
668..679
/misc_feature
/note="encodes nuclear localization signal"
1439..2023
/misc_feature
/note="encodes ring finger"
2619
/misc_feature
/note="alternative splice at nucleotide 2619 truncates
protein before leucine zipper"
BASE COUNT 943 a 773 c 739 g 899 t
ORIGIN

```

Query Match 20.9%; Score 323.6; DB 91; Length 3354;
 Best Local Similarity 59.4%; Pred. No. 1.9e-50;
 Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

```

RESULT 9
HSU26710 3982 bp mRNA PRI 30-SEP-1996
LOCUS Human cbl-b mRNA, complete cds.
DEFINITION U26710
ACCESSION U26710
VERSION U26710.1 GI:862406
KEYWORDS

```

```

QY 363 caggcgacagctggccaaagctggccaatctcttcaaccacatgcacgcagagctgcacgc 422
|||||
Db 739 CAGACGAAATCTCACAAACTGTCCCTTATCTTCATCATCATCTGCGAGAAATCAAGC 798
|||||
QY 423 actcttcccccggggcaaaagtactgtgtgacacatgtaccagctcaccaaggcccccccca 482
|||||
Db 799 AATCTTCCCAATGGTCAATTCAGGGAGATAACTTTCGTATCACAAAAGCAGATGCTGC 858
|||||
QY 483 cacctttgagggaaagtgtggagcccggtgtgtgtgcccgtgggtgagttttagtc 542
|||||
Db 859 TGAATCTCGAGAAAGTTTTTTTGGGACAAAACATATCGTACCAATGAAAGATATTCAGACA 918
|||||
QY 543 cctctggggcacctggccacccctggaaaccaggtgcacagccctggccttggccaccac 602
|||||
Db 919 GTGCCTTCAT3AGGTCCACCAGATTAGCTTAGCTTGGAGCAATGGCTCTAAATCAAC 978
|||||
QY 603 cattgaacctcacctggagcgggcaagtggtccattgttgagttgagcgttctccaggct 662
|||||
Db 979 AATTGATTTAACTTGCATGATTACATTTCAGTTTGAATTTGATATTTTACCAGGCT 1038
|||||
QY 663 ctctcagccctggccaaacactcctcaagaactggcagctcctggcagctcaaccacagg 722
|||||
Db 1039 GTTTCAGCCCTTGGGCTCTATTTCGGGAATTGGAATTTTACCTGTGACATCCAGG 1098
|||||
QY 723 ctacatggccttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 782
|||||
Db 1099 TTACATGGCAATTTCTCATATATGATGAAGTTAAAGCAGGACTACAGAAATATAGCAACA 1158
|||||
QY 783 gccaggcagttacatcttcggccacgtactcgcctgggagcagtgggcctcagcta 842
|||||
Db 1159 ACCCGAAGCTATATTTTCGGGTTAAGTTGACCTCGATTGGGACAGTGGGCCATTGGCTA 1218
|||||
QY 843 tgtgagctcagatggcagcactcctgcagacactcctgcacacacacacacacacacacac 902
|||||
Db 1219 TGTGAC1GGGAGTGGGAATATTTACAGACCATACCTCATGACCATATAAAAGTTACACAGGAACA 1278
|||||
QY 903 gctcctggaggagcagagcggcctcctcctcctcctcctcctcctcctcctcctcctcctc 962
|||||
Db 1279 CCTGATTGATGGCAGCAGGGAAGGATTTTATCTTTATCTCTGATGGGAGAGTTATAATCC 1338
|||||
QY 963 agacctgactgagctcggccaggcagacacccagcagcagcagcagcagcagcagcagcagca 1022
|||||
Db 1339 TGATTTAACTGGATTATGTGAACCTACACCTCATGACCATATAAAAGTTACACAGGAACA 1398
|||||
QY 1023 gctcagctcctcactgggcccattggactccacatttgactcctgcagatctgtgctgagag 1082
|||||
Db 1399 ATATGAATTAATTGTGAATGGGCTCCACTTTTCAGCTCTGTAAAGATTGTGCAGAGAA 1458
|||||
QY 1083 caacaagatgtgaagattcagccgtcgggcactcctcctcctcctcctcctcctcctcctcct 1142
|||||
Db 1459 TGACAAAGATGTCAAGATTGAGCCTTGTGGGCATTTGATGTGCACCTCTTGCCCTTACGGC 1518
|||||
QY 1143 ctggcagcactcggcacagccagacactcctcctcctcctcctcctcctcctcctcctcctcct 1202
|||||
Db 1519 ATGGCAGGATCGGATGTCAGGGCTGCCCTTTGCCCTTTCTGCTGTGTAATAAAAGAACTGA 1578
|||||
QY 1203 ggcctgagttatcacagtttccaggttcaggtcaggtcaggtcaggtcaggtcaggtcaggtc 1262
|||||
Db 1579 GCCATAATCTGGACCCCTTTGATCCAAAGAGATGAAGGCTCCAGGTTCCAGGTTTTCAGCATCAT 1638
|||||
QY 1263 tgaccaggaaggcaggaggttg 1284
|||||
Db 1639 TGACCCCTTTGGCATGCCGATG 1660

```



```

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 2808)
Langdon, W.Y., Institute of Medical and
Direct Submission
Submitted (01-JAN-1991) Langdon W.Y., Institute of Medical and
Veterinary, Scienc, IMVS, Division of Human Immunology, Box 14
Rundel Mail Post Office, Adelaide SA 5000, Australia
3 (bases 1 to 2808)
Blaker, T.J., Shapiro, M., Morse, H.C. III, and Langdon, W.Y.
The sequences of the human and mouse c-cbl proto-oncogenes show
v-cbl was generated by a large truncation encompassing a
proline-rich domain and a leucine zipper-like motif
Oncogene 6 (4), 653-657 (1991)
91232862
JOURNAL
MEDLINE
AUTHORS
TITLE
JOURNAL
MEDLINE
AUTHORS
TITLE
Oncogene 7 (4), 757-762 (1992)
9228506
Location/Qualifiers
1..2808
/organism="Mus musculus"
/db_xref="taxon:10090"
1..2808
/organism="c-cbl"
/evidence="experimental"
1..2808
/organism="c-cbl"
97..2787
/codon_start=1
/product="c-cbl protein"
/protein_id="CAA40394.1"
/db_xref="GI:50315"
/db_xref="SWISS-PROT:P22682"
/translation="MAGNVKSSGAGGGGSGAGGLIGLMKDAFOPHHHHHLSPH
PCVDRKMEKVKMLKMDKVRQCNPNVAKNSPPYILDLPTIOHRLVLSRVEG
KMETLIGENPFYFEMENLMKTKOTISLFKEGKERMEYENSQPRNLTSLIFSHML
AELGKIPSPFOGDTFRITKADAEFRKFAKTEKIVPKWSFRQALHEVPISSGLE
AWAKSLTDLNCDYISVFEDIFTRLFQPMWSLLRNWNSLAVTHPCYMAFLAYDEVK
ARLQKHFKPSYIFRSCITRLGWAIGYVADGNILQTIPIHNNKPLFQALIDGREF
YLFDFGRNCPDLGLCEPTPDHDKVTQCAENDKDKVLEPCGHLMTCSLISWSQS
BQGPCPCRCIKETPIVDPDFRGSSLLRQGAEGAPSPNYDDDDDERADSLFM
MKELAGKVRSPSPFMAPOALPPVPRLLDLQAPVASTSVLGTAKAAGSL
HKDPLPTPLRLPPLPPDRPYSVGAETRPQRPLPCTPGCPDRKLPPVPSR
PGDWSLSTIFKVPVAPNPGDPWNGELNRSILPFLSPQMEPRADVPRLGSLFL
DTMTNMSVPYGPESHPKIKPSSANALYSLAARDLPMPKLPPEQSGSEEDTEYM
TPTSRPVGVQPEKRELEATQSRACDCDQIDSCYEAMKTIQSGALSVAENSAGS
EGNLATATSGPSEENEDGIDYPPVPVAVLAKRLISDISNASSFGLSLDGP
TNFNGSOVPPRPKPPFRINRSKASSYQGGGATANPVATAPSPQLSSEIERLMS
OQSYQDIQKALVIAHNNIEMAKNILEFVSISSPAHVAT"
BASE COUNT 691 a 796 c 693 g 628 t
ORIGIN

Query Match 17.5%; Score 270.6; DB 12; Length 2808;
Best Local Similarity 57.8%; Pred. No. 1.2e-40;
Matches 535; Conservative 0; Mismatches 339; Indels 51; Gaps 1;

QY 303 gctgcctcccgggcggaagtgccacgacgagctcttccggggcggtccagact 362
Db 471 GCAGACTATCAGCCTTTCAAGGAGGGAAGAAGAGGATGTATGAGGAAATCCCGCC 530
QY 363 caggcgacagctggcgaagctggtccatcttccagccacatgcacgacgactgcagc 422
Db 531 TAGCGCAAACTGACCAAAATATCCCTGATCTTCAGCCACATGCTGGCAGAACTGAAAGG 590
QY 423 actcttcccggggcaagtaactgtgacacatgtaccagctcaccacggcccgccca 482
Db 591 CATCTTCGACGCGACTCTTCAAGGAGACACTTTCGGATACTATAAGCGTATGCTGCG 650
QY 483 cactcttgagggaagtgcggagcccggtgtgtgctgcccctgggctgagtttgagtc 542

```

```

Db 651 CGAATTTTGGAGAAAACCTTTTGGTGAAAAGACGATAGTCCCGTGGAAAGACTTTCGACA 710
QY 543 cctcctggggcaccctgcacccctgtggaaccaggctgcacagccctggccttgccaccac 602
Db 711 GGCCCTGTCATGAAGTGCATCCCATCAGTTCTGGCTGGAGGCATGGCTCTGAAGTCCAC 770
QY 503 cattgacctcactgcagggggcaccgtgctccatcttcgagttcgaagcttccacagct 662
Db 771 TATTGATCTGACCTCAATGATTATATTCTTGTCTTTGAATTTGATTTTACACGGT 430
QY 663 ctttcagccatggccacactcctcaagaactggcagctcctggcagtgcaaccacccagg 722
Db 831 GTTTCAGCCCTGCTCTCTTTGCTCAGAAATTTGAAACAGCCTTGTCTAACTCACCTGG 890
QY 723 ctacatggccttctcactcactatgatgaggtccaaaggctctgcagggcctgcaggagcaa 782
Db 891 TTACATGGCTTTCTGACATACGATGAAGTGAAGGCGCCTGCAGAAAGTTCATCCACAA 950
QY 783 gcaggcaggttacatcttcgcggcccgagctgactcctcctggggcagtgggccatcgcta 842
Db 951 ACTGGCAGTATACATCTTTCGGCTGAGCTACTCGTTTGGGTCACTGGGCTATTGGGTA 1010
QY 843 tgtgagctcagatggcagatcctgcagaccatcctctgcaacaaacccctgtcccaggt 902
Db 1011 TGTACTGCCGATGGGAACATTTCTGCAGACATCCGACACAATAAACCCTCTTCCAAGC 1070
QY 903 gctcctggagggacagagagcgtcttacctctaccagatgggaagacccacacccc 952
Db 1071 ACTGATTGATGGCTCAGGAAGGCTTCTATTGTTTCTGTATGGAGAAATCAAAATCC 1130
QY 963 agactcagctgagctggcggcaggcagaccccgagcgcgcctcagctgtcagaggagca 1022
Db 1131 TGACCTGACAGGTTTATGTCAACCACTCTCAAGATCATATCAAAAGTAAC----- 1181
QY 1023 gctgcagctcactggggccatggactccacattgagctcgaacatctgtcagag 1082
Db 1182 -----CCAGATATGTCTCGAGAA 1199
QY 1083 caacaagatgtgaagattgagcgtgctggcgccactgctgcagctgctgcctggctgc 1142
Db 1200 TGATAAGGATGTGAAGATTGAGCCCTGTGGACACCTCATGTGCACATCGCTCGCTCAGTC 1259
QY 1143 ctggcagcactcggacacgcagaccctgcctcctcctgcctgcgagatcaagggctggga 1202
Db 1260 GTGCAGGAATCAAGAGGTGAGGCTGCTCTTTTCCGAGTGTAAATCAAGGTAAGTACTGA 1319
QY 1203 ggcctgagctatctaccagttccac 1227
Db 1320 GCCCATCGTGGTGGATCCGTTTGAC 1344

RESULT 11
CASNS1
LOCUS
DEFINITION
Mouse Cas NS-1 retrovirus gag-onc fusion protein (v-cbl) gene, 3'
end.
ACCESSION
VERSION
J04169.1 GI:323269
KEYWORDS
SOURCE
gag-onc fusion protein; viral oncogene.
CAS NS-1 retrovirus DNA, isolated from mouse (strain NFS/N)
previously infected with Cas-Br-M virus.
ORGANISM
Cas NS-1 murine leukemia virus
Viruses; Retroviral viruses; Retroviridae; Mammalian type C
retroviruses; 1-Mammalian type C virus group.
REFERENCE
1 (bases 1 to 2711)
Langdon, W.Y., Hartley, J.W., Klincken, S.P., Ruscelli, S.K. and
Morse, H.C. III.
TITLE
v-cbl, an oncogene from a dual-recombinant murine retrovirus that
induces early B-lineage lymphomas
Proc. Natl. Acad. Sci. U.S.A. 86, 1168-1172 (1989)
89145204
JOURNAL
MEDLINE
COMMENT
Draft entry and computer-readable sequence for [1] kindly submitted
by W.Y. Langdon, 02-MAR-1989. Cas NS-1 is an acutely transforming

```


RESULT	13
CEDNASLI1	
LOCUS	CEDNASLI1 2230 bp DNA
DEFINITION	C.elegans DNA for sli-1 gene.
ACCESSION	X89223
VERSION	X89223.1 GI:895766
KEYWORDS	negative regulator; sli-1 gene.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis
REFERENCE	1 (bases 1 to 2230)
AUTHORS	Yoon, C.H., Lee, J., Jongeward, G.D. and Sternberg, P.W.

QY 782 agccaggcaggtatcatcttcgagccacagctgtactgcctgggagcagtgaggccatcgct 841
 DB 951 AACCTGGAGAGTACATCTCCGGTTATCATGCACAGCTCTGGCAATGGGCAATAGGAT 1010
 QY 842 atgtgagctcagatggcagcagctctcagagccatccctgccacacacccctctccacg 901
 DB 1011 ACGTAGCTCCGGTGAAGAAGATTATCAGACAATACCACAGAAATAAAAGTTTGATTCAG 1070
 QY 902 tgcctctgagggagcagagagcagctctcactctcaccagatggagaaagccacacacc 961
 DB 1071 CACTACATGAAGGCCATAAAGAAGATTTATATTTACCGCAACGGTAGAGATCAAGATA 1130
 QY 962 cagacctgactgagctgcccagcagagaccccccagcagcagctccagctgcagagagc 1021
 DB 1131 TTAATTATCCAAATTGATGATGTGCCACAAAGCGGACAGTCAAGTACACAGTGAAC 1190
 QY 1022 agctgcagctctactgggcccagtgactccacattgtgactctcgaagatctgtgctgaga 1081
 DB 1191 AATACGAGTTCTGATGTGAGATGGGCACAACTTCGAGTTGTGCAAAATTTGTGACGATA 1250
 QY 1082 gcaacaagatgtgaagattgagcgtgagcggcagctctcagagctgtgctgctgctg 1141
 DB 1251 ACGAGAGAACATCAAAATGAGCCATGTGACATTTGCTCTGGCAGAAATGTTGCTA 1310
 QY 1142 cctggcagcactcgga-----cagccagacctgcccctcttgcgcgtgcgagatcaag 1195
 DB 1311 ACTGGCAGGATTCGATGTTGGTGGCAACACATGTCATTCTGCGCGTACGAATCAAG 1370
 QY 1196 gctgggagcgtgagtatctaccagttcca 1226
 DB 1371 GAACAATCTGTGATTATTCACACAGTTCAA 1401

RESULT 14

AC014994/c

LOCUS

DEFINITION Drosophila melanogaster, ** SEQUENCING IN PROGRESS **, in ordered

pieces.

ACCESSION AC014994

VERSION AC014994.1 GI:6436341

KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 36890)

AUTHORS Adams, M. and Venter, J. C.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

COMMENT This sequence was identified as CDN:10212090 by the submitter.

For further information on this sequence e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

1..36890

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 11572 a 7373 c 7329 g 10616 t

ORIGIN

Query Match 11.6%; Score 179.2; DB 41; Length 36890;

Best Local Similarity 65.5%; Pred. No. 4e-24;

Matches 262; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 518 tgcctgcccctgggctgagttgagctcctctgagcagctccacccctgggagcagact 577

DB 30388 TGGTTCCTCGGAAATCTCCGCGCAGGAGCTTAACAAGTACATCCATPAATCTCCGGCC 30329

QY 578 gcacagccctgagccttgccgacacacacattgacacctgcagcgggacagtggtccatct 637
 DB 30328 TGGAGCCCATGGCCCTAAAGACCACTATCGATCTTACCTGCAACGACTTCTTCCAACT 30269
 QY 638 tgcagttcagcctctccacagcgtctcttcagccatggccacacactcctcaagaactggc 597
 DB 30268 TCGAGTTCCAGCTCTTCACACGCCCTTCCAGCCCTTGGTGACCTGCTACGCAACTGGC 30209
 QY 698 agctcctggcagctcaacacacacacagcagctacatggcctcctcaccctatgatgaggtcccaag 757
 DB 30208 AGATTTCGCCCTCACACATCCGGCTACGTGGCGTTTCTTCACATACGACGAGTGAAGG 30149
 QY 758 acggtctgagcgtctcagggagacacagcagcaggttacattcttcggccagcagtgtaact 817
 DB 30148 CTCGGCTACAGCGCTACATCTCAAGCGGCGACGCTACGTTCCTCGGCTCTCTCGACGC 30089
 QY 818 gctctgggagcagtgaggccatcgctatgtgagtcagatggcagcagcctcctgcagaccatcc 877
 DB 30088 GATTGGCCCAATGGCCATCGCTACGTAACTGCGGAGGAGAGATCTCGCAGACATCC 30029
 QY 878 ctggcaacaacacccctgtccaggtgctcctggaggagaca 917
 DB 30028 CTGAGAACAACTCGCTGTGCCAGCGCTGCTCGATGGCCA 29989

RESULT 15

LOCUS

AC010033

DEFINITION

Drosophila melanogaster clone RPI98-315, ** SEQUENCING IN

PROGRESS:***, 75 unordered pieces.

ACCESSION AC010033 AC023749

VERSION AC010033.7 GI:7958624

KEYWORDS HTG; HTGS_PHASE1.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 129226)

AUTHORS Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,

Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunick, C.,

Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, L., Cox, C.,

David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,

Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,

Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,

Guevara, W., Harris, K., Hernandez, J., Hodgeson, A., Hoques, M.,

Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,

Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,

Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,

Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,

Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,

Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,

Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,

Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,

Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wabhang, M.,

Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,

Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and

Gibbs, R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 129226)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (11-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 20, 2000 this sequence version replaced gi:5996835

gi:5996719.

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Fri Dec 8 06:14:02 2000

```
----- Project Information
Center project name: DRGJ
Center clone name: RPC198-315
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 62108 bases at least Q40
Consensus quality: 84593 bases at least Q30
Consensus quality: 95829 bases at least Q20
Estimated insert size: 100104; sum-of-contigs estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1      8568: contig of 8568 bp in length
      8668: gap of unknown length
      8669      8669: contig of 4728 bp in length
      13397      13396: contig of 4200 bp in length
      13497      17696: contig of 4200 bp in length
      17697      17996: gap of unknown length
      21861      21860: contig of 4064 bp in length
      21961      21960: gap of unknown length
      25099      25098: contig of 3138 bp in length
      25199      25198: gap of unknown length
      27961      27960: contig of 2762 bp in length
      28061      28060: gap of unknown length
      30797      30796: contig of 2736 bp in length
      30897      30896: gap of unknown length
      33632      33631: contig of 2735 bp in length
      33732      33731: gap of unknown length
      36353      36352: contig of 2622 bp in length
      36354      36353: gap of unknown length
      36454      36453: gap of unknown length
      39019      39018: contig of 2565 bp in length
      41507      41506: gap of unknown length
      41607      41606: gap of unknown length
      43946      43945: contig of 2339 bp in length
      44046      44045: gap of unknown length
      46295      46294: contig of 2249 bp in length
      46395      46394: gap of unknown length
      48396      48395: gap of unknown length
      48593      48592: contig of 2198 bp in length
      48694      48693: gap of unknown length
      50813      50812: contig of 2120 bp in length
      50914      50913: gap of unknown length
      53017      53016: contig of 2104 bp in length
      53117      53116: gap of unknown length
      55095      55094: contig of 1978 bp in length
      55195      55194: gap of unknown length
      55196      55195: contig of 1960 bp in length
      57156      57155: gap of unknown length
      57255      57254: gap of unknown length
      59209      59208: contig of 1954 bp in length
      59310      59309: gap of unknown length
      61213      61212: contig of 1904 bp in length
      63167      63166: gap of unknown length
      63267      63266: gap of unknown length
      65097      65096: contig of 1830 bp in length
      65197      65196: gap of unknown length
      66300      66299: contig of 1733 bp in length
      67030      67029: gap of unknown length
      68684      68683: contig of 1654 bp in length
      70393      70392: contig of 1609 bp in length
      70493      70492: gap of unknown length
      72081      72080: contig of 1588 bp in length
      72181      72180: gap of unknown length
      72082
```

```
72182      73749: contig of 1568 bp in length
73750      73849: gap of unknown length
73850      75387: contig of 1538 bp in length
75388      75487: gap of unknown length
75488      77000: contig of 1513 bp in length
77001      77100: gap of unknown length
77101      78500: contig of 1500 bp in length
78501      78700: gap of unknown length
78701      80199: contig of 1499 bp in length
80199      80299: gap of unknown length
80300      81739: contig of 1440 bp in length
81740      81839: gap of unknown length
81840      83368: contig of 1429 bp in length
83369      83368: gap of unknown length
83369      84932: contig of 1424 bp in length
84933      84932: gap of unknown length
84934      86250: contig of 1358 bp in length
86251      86350: gap of unknown length
86351      86350: gap of unknown length
86351      87676: contig of 1326 bp in length
87677      87776: gap of unknown length
87777      89094: contig of 1318 bp in length
89095      89194: gap of unknown length
89195      90449: contig of 1255 bp in length
90450      90549: gap of unknown length
90550      91801: contig of 1252 bp in length
91802      91901: gap of unknown length
91903      93127: contig of 1226 bp in length
93128      93227: gap of unknown length
93228      94440: contig of 1213 bp in length
94441      94540: gap of unknown length
94541      95730: contig of 1190 bp in length
95731      95830: gap of unknown length
95831      97015: contig of 1185 bp in length
97016      97115: gap of unknown length
97116      98210: contig of 1095 bp in length
98211      98310: gap of unknown length
98311      99370: contig of 1060 bp in length
99371      99470: gap of unknown length
99471      100529: contig of 1059 bp in length
100530      100529: gap of unknown length
100531      101680: contig of 1051 bp in length
101681      101780: gap of unknown length
101781      102843: contig of 1045 bp in length
102844      102925: gap of unknown length
102926      103948: contig of 1023 bp in length
103949      104048: gap of unknown length
104049      105054: contig of 1006 bp in length
105055      105154: gap of unknown length
105155      106133: contig of 999 bp in length
106134      106253: gap of unknown length
106254      107244: contig of 991 bp in length
107245      107344: gap of unknown length
107345      108328: contig of 984 bp in length
108329      108428: gap of unknown length
108429      109408: contig of 980 bp in length
109409      109508: gap of unknown length
109509      110479: contig of 971 bp in length
110480      110579: gap of unknown length
110580      111536: contig of 957 bp in length
111537      111636: gap of unknown length
111637      112557: contig of 921 bp in length
112558      112657: gap of unknown length
112659      113571: contig of 914 bp in length
113572      113671: gap of unknown length
113673      114559: contig of 888 bp in length
114560      114659: gap of unknown length
114660      115542: contig of 883 bp in length
115543      115642: gap of unknown length
115643      116515: contig of 874 bp in length
116516      116616: gap of unknown length
116617      117461: contig of 845 bp in length
117462      117561: gap of unknown length
117562      118402: contig of 841 bp in length
118403
```

```

* 118403 118502: gap of unknown length
* 118503 119341: contig of 839 bp in length
* 119342 119441: gap of unknown length
* 119442 120279: contig of 838 bp in length
* 120280 120379: gap of unknown length
* 120380 121210: contig of 831 bp in length
* 121211 121310: gap of unknown length
* 121311 122141: contig of 831 bp in length

Query Match
Best Local Similarity 11.6%; Score 179.2; DB 39; Length 129226;
Matches 262; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 518 tgcctgctggctgagtttgcctcctggtggaactgcccacccctgtggaaccaggct 577
Db 122325 TGGTTCCTGAAAAATCTCCGGCAGGAGCTTAACAAAGTACATCCCAATCTCCGGCC 122384

QY 578 gcacagccctggccttcgcacacacattgaactcaactgcagcgggcaagtgtccatct 637
Db 122385 TGGAGGCCATGGGCCATAAGACCACCTATCGATCTTACCTGCAACGACTTCATTCCAACT 122444

QY 638 tcgagttcgacgtttccaccaggctcttcagccatggcccaacactccctcaagaactggc 697
Db 122445 TCGAGTTTCGAGCTTTCACACGCTCTTCCAGCCTTGGGTGACCCCTGCTACGCAACTGGC 122504

QY 698 agctcctggcagtcacacacacccagggtacatggccttcctcaactatgatgaggtccaa 757
Db 122505 AGATTCTGGCCGTACACATCCGGGCTACGTGGCGTTTCTACATACGACGAGGTGAAGG 122564

QY 758 agcgtctgcagcctgcagggacaagccaggcagttacatcttcggcccgccagctgtactc 817
Db 122565 CTGCGCTACAGCGCTACATCCTCARGCGGGCAGCTACGTTTCCGGGCTCTCTGACGCG 122624

QY 818 gcctggggcagtgggccatcgggtatgtgagctcagatggcagcatcctgcagaccatcc 877
Db 122625 GATTGGGCCAATGGGCATCGGCTACGTAACTGCCGAGGGAGAGATTCTGCAGACAATCC 122684

QY 878 ctgccaacaacccctgtcccgagtgctcctcgaggagaca 917
Db 122685 CTCAGAACAAAGTCGCTGTGCCAGGCGCTGCTGATGGCCA 122724

```

Search completed: December 7, 2000, 13:43:21
Job time: 27733 sec